

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 01:21:45 : Search time 3118 Seconds
(without alignments)
5684.288 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMIRPTALLPALFFPHAY.....IAYGHSKKIRWHISLGTRF 609

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+p2n.model -DB=us09-857-669-2
-Q=/cgn2_1/USPTO.spool/US09857669/runat_05112002_110007_7724/app_query.fasta_1.775
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669@cgn_1.1.1879 -runat_05112002_110007_7724 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.un.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	3173	100.0	1830	6	AX026688 Sequence
2	3173	100.0	1830	6	AX026690 Sequence
C 3	3153	99.4	340806	1	NMA122491
4	3136	98.8	12429	1	AE002561 Neisseria
5	3136	98.8	172325	6	AX044035 Sequence
6	3019	95.1	1764	6	AX026692 Sequence
C 7	596	18.8	189050	1	AL646069
C 8	449	14.2	11158	1	AE004682
9	384.5	12.1	1374	6	AX078576 Sequence
10	380.5	12.0	243184	1	AP002568 Escherich
11	379.5	12.0	10819	1	AE000493 Escherich
12	379.5	12.0	11760	1	AE005634 Escherich
13	379.5	12.0	338534	1	EC00W93
14	374	11.8	1737	6	AX033468 Sequence
C 15	374	11.8	11833	1	U32752
16	366.5	11.6	10831	1	AE013668 Versinia
C 17	366.5	11.6	216050	1	AJ414157
18	357	11.3	24892	1	AE008906
19	357	11.3	249050	1	AL627283
20	355.5	11.2	1731	6	AX033470 Sequence
C 21	348	11.0	14390	1	AE004323
22	342.5	10.8	201050	1	AL646064
23	342	10.8	333800	1	SME591792
C 24	336	10.6	10302	1	AE006218
25	323.5	10.2	10029	1	AE008175
26	323.5	10.2	10301	1	AE009208
27	306.5	9.7	10074	1	AE003957
C 28	304.5	9.6	10029	1	AE009623
C 29	300	9.5	39013	1	AF088896
30	292	9.2	20310	1	AY010120
C 31	291.5	9.2	10839	1	AE012071
C 32	291	9.2	11870	1	AE012531
33	283	8.9	334520	1	AP003588
34	272.5	8.6	10301	1	AE005835
35	260.5	8.2	2379	1	NGU81959
36	256.5	8.1	2379	6	AX155443
37	256.5	8.1	2379	6	AX155447
38	253.5	8.0	2394	6	AX155447
39	253.5	8.0	2394	6	AX202497
40	253.5	8.0	340806	1	NMA122491
C 41	253.5	8.0	347750	1	AP002998
42	253	8.0	2394	6	AX155439
43	253	8.0	2394	6	AX202489
C 44	253	8.0	13538	1	AE002375
C 45	253	8.0	349980	6	AX043922

ALIGNMENTS

RESULT 1

AX026688
 LOCUS AX026688 1830 bp DNA linear PAT 16-SEP-2000
 DEFINITION Sequence 1 from Patent WO0034480.
 ACCESSION AX026688
 VERSION AX026688.1 GI:10187860
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis.
 Neisseria meningitidis
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 1 (bases 1 to 1830)
 Ruelle, J.-L.
 Polypeptides and polynucleotides basb040 from neisseria
 meningitidis and vaccine comprising said polypeptides and
 polynucleotides
 Patent: WO 0034480-A 1 15-JUN-2000;
 RUELLE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
 JOURNAL
 FEATURES
 source
 Location/Qualifiers
 1..1830
 /organism="Neisseria meningitidis"
 /db_xref="taxon:487"
 BASE COUNT 440 a 603 c 461 g 326 t
 ORIGIN

Alignment Scores:

Pred. No.: 1..17e-204 Length: 1830
 Score: 3173.00 Matches: 609
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-857-669-2 (1-609) x AX026688 (1-1830)

Qy 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
 Db 1 ATGATGATCAACCGGACCGGCTGCTCTCGCGGCTTTATTTTCTTCGCGACGCATAC 60
 Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
 Db 61 GCGCTCCCGCCGACCTTCCGAAACACAGCGCGGCTTTCGATTTGTTCAAAAACAAA 120
 Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
 Db 121 AGCCCCGACACCGAATCAGTAAATTAATTAACCAACCAATTCGCGTCCGATCGACACG 180
 Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
 Db 181 GATAGTAAATCAAGATATGGTGAAGAACACCTGCGGCTCATCGCGACGACAGGAA 240
 Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
 Db 241 GAAGTATGGACAAGGAACAGACGGGCTTCTCCCGAAGACGACCGGACACGTTAAA 300
 Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
 Db 301 ACCATGCTCCGCGACCAAGGCTATTTTCAGCAGCAAACTCAGCCTGACGGAAAAAGACGGA 360
 Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
 Db 361 GCTTATACGGGTACACATACACCGGGGGCCGCGACCAAAATCGCCACGTCGGTGTCCGC 420
 Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
 Db 421 ATCCCTCGCGACATCCTTTTCAGAGGCAACCTCGCCGAATACTACCGACGCGCTGAA 480
 Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
 Db 481 AACTGGCAGCAGCGGTAGGACGTGATTTCGATCAGACACAGTTGGGAAAAACAGCAAACT 540
 Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
 Db 541 TCCGTCTCGCGCGGTAAACGCAAAAGGCTACCCGCTTCCCAAGCTTCGCGCAACACCCGG 600

Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
 Db 601 GCGCGCTCAACCCCGATACCGCACCGTCGATTGAACCTCGTTCGTGGACAGCGGCGCG 660
 Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
 Db 661 CCCATCGCTTCGCGGACTTTGAAATCAGCGGACACAGCGTTACCCGCAACAAATCGTC 720
 Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuLeuLeuLeuAspPhe 260
 Db 721 TCGCGCTCGCGGCTTCCAAACCGGCGACGCCCTACGACCTCGACCTGCTCGACTTC 780
 Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
 Db 781 CAACAGCCTCGAACAACAAACGCGCATTTATTCGCGCGCTCCGTACAAAGCGGACTTCGAC 840
 Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
 Db 841 CGTCTCAAGGCGACCGCTCCCGCTCAAAAGTCAGCGTAACCGAGGTCAAAACGCCACAAG 900
 Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
 Db 901 CTCGAAACCGCGCATCCGCTCGATTCGGAATACGGTTTGGGCGCAAAATCGGCTACGAC 960
 Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
 Db 961 TATTACAACTCTTCAACAAAGGCTATATCGGCTCGGCTCGGATATGGACAAATAC 1020
 Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 Db 1021 GAAACCAAGCTTTCGCGCGCATCAGCAGCGCGCAACTATCGGGCAACTACTTGACAA 1080
 Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
 Db 1081 AGCAACGTTTCCATCAACACCGTTCGACCAACCAACCGTTCGAAACCGCGCTTCTCGCG 1140
 Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
 Db 1141 GGCATCTGGTATGTGCGGACCGCGCGGCGATCGATGCCAGGCTGGGGGCGAGTTCTC 1200
 Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
 Db 1201 GCAGAGCGCGGAAATCCCGGCTCGGATATCGATTTCGGCAACACGCCACGCCACGATG 1260
 Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
 Db 1261 CTGACCGCTTCTGGAAACCGCGCTGCTCAACACGCTGCTGCTCCGAAACGCGCAT 1320
 Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
 Db 1321 TACCTCGACGCAAAATCGGTACGACTTTGGGCACATTCCTGCTCCACCGCGCTGATC 1380
 Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
 Db 1381 CGCACCTCTCCCGTGCAGGTATTTCTTCCGCGGCTTACGAAACCAAAACAACTCGCACGCTTC 1440
 Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAlaAspValProSerGly 500
 Db 1441 ATCATCGCGGACAGCGGCTTACCGTTGCGCGGACAAATCGCGACGTTCTTCAGGG 1500
 Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
 Db 1501 CTGATGTCGCGACGCGCGGCTCTTCTCGTGGCGGTTACGAACTCGACAGCATCGGA 1560
 Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
 Db 1561 CTTCCGCGCGGCAACGATCGGCTCTGCCGAAACGCGCGCTCTCTGTTGGGCGAGCTGAA 1620
 Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyValAlaValPheHisAspMetGlyAspAla 560
 Db 1621 TACCACTCGCGGTTTACGCGCACCTTTCGCGCGGCTGTTTCCACGATATGGGCGATGCC 1680

```
QY 561 AlaAlaAsnPhyArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 1681 GCCGCAATTTCAAAGCTATGAAGTGAACACACGGTTCGGGACTGGCGCTGGCTGGTTTC 1740
QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 1741 AGCCCGCTTGCCTGCTTTCCTTCGACATCGCTACGGGCACAGGATAGAAGAAATCCGC 1800
QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 1801 TGGCACAATCAGCTTGGGAACACGCTTC 1827

RESULT 2
AX026690
LOCUS AX026690 1830 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 3 from Patent WO0034480.
ACCESSION AX026690
VERSION AX026690.1 GI:10187861
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 1830)
AUTHORS Ruelle,J.L.
TITLE Polypeptides and polynucleotides basb040 from neisseria
meningitidis and vaccine comprising said polypeptides and
polynucleotides
JOURNAL Patent: WO 0034480-A 3 15-JUN-2000;
RUELLE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
Location/Qualifiers
1..1830
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 439 a 604 c 461 g 326 t
ORIGIN
Alignment Scores:
Pred. No.: 1,17e-204 Length: 1830
Score: 3173.00 Matches: 609
Percent Similarity: 100.00% Conservatative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-857-669-2 (1-609) x AX026690 (1-1830)
QY 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 1 ATGATGATCAAAACCGACCGCTGCTGCTGCGGGCTTATTTTCTTCCTCCGACGCATAC 60
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 61 GGGCTTGGCCGCGACCTTTCCGAAAAACAAGGGCGGGTTTCGATTTGTCAAAAACAA 120
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 121 AGCCCGACACCAATCAGTAAATTAACCAAAATTCCTCCGCTCGCATCGACACGAG 180
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln 80
Db 181 GATAGTGAATCAAGATATGGTGAAGAACACCTCGCGCTCATCACGACGACGAGAA 240
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 241 GAAGTATGGACAAGGAACAGACGGCTTCCTCGCGGAAGACACCGGACACGTTAA 300
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 301 ACCATGCTCCGACGAAGAGGCTATTTACAGCAAGTACAGCTGACGGGAAGACGGA 360
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 1441 ATCATACGGGACAAAGGGGTTACACCGTTGCCCGGACAAATGCCGACGTTCTTCCAGG 1500
```

```
Db 361 GCTTATACGGTACATACACACCGGCGCGCCGCAACAAATCGCCAAATCGGTGTGTGCC 420
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 421 ATCCCTCGGCGACATCTTTTCAGACGGCAACCTCGCGGAATACTACCGCAACGCGTGGAA 480
QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 481 AACTGGCAGCAGCGGTAGCAGTGTATTCGATCAGGACAGTTGGGAAACAGCAAACT 540
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 541 TCCGCTCTCGGCGGTAAACGCAAGGCTACCGCTTCCCAAGCTCGCAACACCGCG 600
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 601 GCGGCGGTCAACCCGATACCGCCACCGCTGATTTGAACGTGCTGGACAGCGCGCG 660
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 661 CCCATCGCTTCGGCGACTTTGAATCATCCGGCACACACAGGTTACCCCAACAAATCGTC 720
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 721 TCCGCGCTGGCGGCTTCCAAACGGGCGCGCTACGACCTCGACCTGCTGCTCGACTTC 780
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 781 CAACAGGCACTCGAACAAACGGGCATTATTCGGCGCTCGGTACAAACGCGACTTCGAC 840
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 841 CGTCTCCAAGCGACCGCGCTCCCGCTCAAGTCAGGTACCGGAGTCAACCGCCACAG 900
QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
Db 901 CTCGAAACCGGATCGCGCTCGATTCGGAATACGGTTTGGGGCGCAAAATCGCTACGAC 960
QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 961 TATTACAACCTCTTCAACAAAGGCTATATCGGCTCGGTCTGCGGATATGGCAAAATAC 1020
QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 1021 GAAACACCGCTTCCCGCGGATCATCCAGCGCGCACTATCTCGGGCACTACTCGACA 1080
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 1081 AGCAACGTTTCTTACAAACGTTTCGACCAACCAAAACCTCGAAAAACGCGCTTCTCCGGC 1140
QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 1141 GGCATCTGGTATGTGCGGACCGCGGCGCATCGATGCCAGGCTGGGGCGAGATTCTC 1200
QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 1201 GCAGAGCGCGGAAATCCCGGCTCGGATATCGATTGGGCAACAGCCACGCGCAGATG 1260
QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
Db 1261 CTGACCGCTCTTGGAAAGCGGAGCTGCTCAACACAGTGTGTCATCCGAAAAACGGCCAT 1320
QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 1321 TACCTCGACGGCAAAATCGGTACGACTTTGGGCAACATTCCTGCTCCACCGCGCTGATC 1380
QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 1381 CGCACCTTCCCGCTCGCGTATTTCTTCCACCGCGGAAACAAACAACTCGGACGCTTC 1440
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 1441 ATCATACGGGACAAAGGGGTTACACCGTTGCCCGGACAAATGCCGACGTTCTTCCAGG 1500
```

gene	<p>complement(414. .2438)</p> <p>/gene="nuol"</p> <p>complement(414. .2438)</p> <p>/gene="nuol"</p> <p>/EC_number="1.6.5.3"</p> <p>/note="NMA0002, nuol, NADH dehydrogenase I chain L, len: 674 aa: similar to many e.g. NUOL_RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored_q1, NADH-ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored_q1_N, NADH-ubiquinone oxidoreductase (complex I), chain 5 N-terminus"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="NADH dehydrogenase I chain L"</p> <p>/protein_id="CAB83322.1"</p> <p>/db_xref="GI:7378780"</p> <p>/db_xref="SWISS-PROT:Q9JX92"</p> <p>/translation="MNDMTYLILALVPLAGSLIAGFNKIGRGAHTVTTLGVAVS AVLSAYVLWGLFNGSLGKADPENVTWLTMGGLDFSVGLFVDTMTAMMVVYTCVSLMV VITIGYMHDEKVGQYRFFSYISLFTFSMLMLINSNFTQLEFGWEAVGLSVYLLIGF YFKRPSFAFLKAFILNRVGDGFLGIGLVLAYFGGSLRGYQDFAYILPNNQVATIQ LFPQVMSLTITVCLLLIFVCGMKSAQFLPHVWLPDSMEGTPISALIIAATWMTAGL FMVSRMSPIVEMSTALSVMTVIGATLPMGFLGVQNDIKRVVAYSTLSQLGYMTV ALGASYSVAMFIVHTAPFKALLFLAASATIGMHDDQMRHGNLKYMPITWLTMT LIGNLSLITPFFSGFYSGKSDIIIEAKSYTLPGSGFAYFAVLASFEVTPAFAPQTFM VFIGEEKRSLPHEHSDGHEHGLKNDNPHESLPLVTLLIILAVSPVITIGYIAI EPMLYGDFDFVADPTVMAHDPTWHIKKEEFHGALAMVSHSLSPVLIYLAAGVLSAWL LYVLPHPAKIAQAPRPVYLVLENKYVILDALYFNVFAKGTALGTFPFWKVGDTAIDD NLGVNSARLVGVAAGVRKQVGTGFIYTYAAMVFGVLVLLGNTFWGLFR"</p> <p>1035. .1044</p> <p>/note="Core DNA uptake sequence: gccgtctgaa"</p> <p>/label=DUS</p> <p>complement(1116. .2030)</p> <p>/gene="nuol"</p> <p>/note="Pfam match to entry PF00361 oxidored_q1, NADH-ubiquinone/plastoquinone (complex I), various chains, score 351.80, E-value 7.3e-102"</p> <p>complement(2061. .2243)</p> <p>/gene="nuol"</p> <p>/note="Pfam match to entry PF00662 oxidored_q1_N, NADH-ubiquinone oxidoreductase (complex I), chain 5 N-terminus, score 77.20, E-value 3.5e-19"</p> <p>complement(2487. .2828)</p> <p>/gene="NMA0003"</p> <p>complement(2487. .2828)</p> <p>/gene="NMA0003"</p> <p>/note="NMA0003, unknown, len: 113 aa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="hypothetical protein NMA0003"</p> <p>/protein_id="CAB83323.1"</p> <p>/db_xref="GI:7378781"</p> <p>/db_xref="SPTREMBL:Q9JX91"</p> <p>/translation="MNDKNLEDEEFHAKMLNIYQELNLPPQYKATRFQLIVNFEFGK FAADKLSTGEKKTQTGTETLLISGGGVHAKYSMEYVLQRPWCDLTFEEQLAVARK RLERGVGFSPK"</p> <p>complement(2833. .2837)</p> <p>complement(2952. .3527)</p> <p>/gene="NMA0004"</p> <p>complement(2952. .3527)</p> <p>/gene="NMA0004"</p> <p>/note="NMA0004, unknown, len: 191 aa: similar to hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa), fasta scores; Et): 0; 80.5% identity in 190 aa overlap, and to proteins involved in cell cycle e.g. TR:O54679 (EMBL:AF036487) Lactococcus lactis plasmid pNZ4000</p>
CDS	
misc_feature	
misc_feature	
misc_feature	
gene	
CDS	
gene	
CDS	
RBS	
gene	
CDS	

putative mobilization protein (200 aa), fasta scores: E():
0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
filamentation protein FIC (200 aa), fasta scores: E():
0.093, 29.4% identity in 126 aa overlap"

/codon_start=1

/transl_table=11

/product="hypothetical protein NMA0004"

/protein_id="CAB83324.1"

/db_xref="GI:7378782"

/db_xref="SPTREMBL:Q9J0R9"

/translation="MPSENPIGKTHKSIDEOSLHNRRLPFGSGDIDRIEVTGTTAGLQQ
IHYRUGGLIFAGQIDEDNISKGGFRANAMYLKALVKIEQMPERTFEIIAIVE
MNIAPFLLEGSRIRIWLDLVLKKNLKKVNVQNSKTYLQAMERSFPVNDLEURL
LKDNITDDVNRRIIFKGIQSYVEGYEK"

complement(3516..3525)

/gene="NMA0004"

/note="Core DNA uptake sequence: gccgtctgaa"

/label=DUS

complement(3561..3866)

/gene="nuoK"

complement(3561..3866)

/gene="nuoK"

/EC_number="1.6.5.3"

/note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase
I chain K (EC 1.6.5.3) (102 aa), fasta scores: E():
1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
match to entry PF00420 oxidored_q2.

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"

/codon_start=1

/transl_table=11

/product="NADH dehydrogenase I chain K"

/protein_id="CAB83325.1"

/db_xref="GI:7378783"

/db_xref="SPTREMBL:Q9J0U9"

/translation="MITLTHYLVLGALLFGISAMGIFMNRKNVLLMSIELMLLAVN
FNFIASQHLGDTAGQIFVFLTVAAAEASAIGLAIVLVYRNROTINVDIDELKG"

complement(3564..3863)

/gene="nuoK"

/note="Pfam match to entry PF00420 oxidored_q2.

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,

score 141.50, E-value 1.5e-38"

complement(3863..4534)

/gene="nuoJ"

complement(3863..4534)

/gene="nuoJ"

/EC_number="1.6.5.3"

/note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fasta scores: E():
9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3.

NADH-ubiquinone/plastoquinone oxidoreductase chain 6"

/codon_start=1

/transl_table=11

/product="NADH dehydrogenase I chain J"

/protein_id="CAB83326.1"

/db_xref="GI:7378784"

/db_xref="SPTREMBL:Q9JX90"

/translation="MTFQILFIYFAVILYGAIKTVTAKNPVHAALHVLVFCVSAM
LMLMQAEFLVTVVYVAVMVLFLFVYMLNIDIEEMRAGFWRHAPVAGVGTLL
AVALLILVNPRTDLAAFGMLKIDIPADYNNIRDLGSRIVTYLLPFLAALVLLGMV
AAIALVHRKTVNPKRMDPADQVKVRADQGRMLVKMEAVKPKQTESAESEVSDDLKPK
EKGKA"

complement(3874..3879)

/gene="nuoJ"

complement(4034..4525)

/gene="nuoJ"

/note="Pfam match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"

complement(4570..5277)

/gene="NMA0007"

complement(4570..5277)
/gene="NMA0007"
/note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"

Alignment Scores:

Pred. NO.: 2.48e-200 Length: 340806
Score: 3153.00 Matches: 605
Percent Similarity: 99.51% Conservative: 1
Best Local Similarity: 99.34% Mismatches: 3
Query Match: 99.37% Indels: 0
DB: 1 Gaps: 0

US-09-857-669-2 (1-609) x NMA1Z2491 (1-340806)

QY 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 289737 ATGATGATTAAACCGACCGCCCTGCTCTCGCGGCTTTATTTTCTTCGCGACGCATAC 289678
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 289677 GCGCCTGCGCGCACCTTTCCGAAAAACAAGGCGCGGTTTCGCAATTGTTCAAAAAACAA 289618
QY 41 SerProAspThrGlnSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 289617 AGCCCGACACCGCAATCAGTTAAATTAACCCCAATTCCTCCGCAATCGACACGCAG 289558
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 289557 GATAGTGAATCAACAGATATGTCGAAGAACACCTCCCGCTCATCAGCAGCAGAGGAA 289498
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 289497 GAAGTATTGGACAAAGAACAGCGGCTTCCTCGCGAAGAACCGGCAACGTTAA 289438
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 289437 ACAATGCTCCGACGCAAGGCTATTTTCAGCAGCAAGATCAGCTGACGGAAGAACGGA 289378
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 289377 GCTTATACGTACATACATCACCGCGCCCGCCGCAACAAATCGCAACGCTCGCGTCC 289318
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 289317 ATCCTCGCGACATCCTTTTCAGACGCGCAACCTCCCGCAATACCGCAACGCGTGGAA 289258
QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 289257 AACTGCGACGACGCGGTAGCAGTGTATTCGATCAGACAGATTTGGGAAAAACAGCAAACT 289198
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 289197 TCGTCTCTCGCGGGTAAACGCGCAAGGCTACCCGCTTCCCAAGCTCGGCAACACCGCG 289138
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 289137 CGCGCGCTCAACCCGATACCGCGGATTTGAACGTCGTCGTCGTCGTCGTCGTCGTCGTCG 289078
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 289077 CCCATCGCCTTCGCGGACTTTGAAATTAACCGCAGCAGCGGCTTACCCCGAACAATCGTC 289018
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 289017 TCGCGCTTGGCGGCTTCCCAACGCGGACCGCCCTTACGACCTCGACCTGCTCGACCTC 288958
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
|||||

Db	288957	CAACAGCGCGTCGACAAACAAAGCGGCATATTTCGCGCGCGTCGCTACAAAGCGCGACTTCGCAC	288898
Qy	281	ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys	300
Db	288997	CGCTCCAGCGCAGCGCGTCCCGCTCAAGTACAGCTACCGAGGTCAACGCCACAAG	288838
Qy	301	LeuGluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGlyLysIleAlaTyrAsp	320
Db	288837	CTCGAAACCGGCATCCGCTCGATTCCGAATACGGTTTGGCGCGCAAAATCGCTACGAC	288778
Qy	321	TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValLysPheMetAspLysTyr	340
Db	288777	TATTACAACTCTTCAACAAAGGCTATATCGGTTCTGCTGGGATATGGACAATAC	288718
Qy	341	GluThrThrLeuAlaLysIleSerGlnProArgAsnTyrArgGlyAsnTyrTyrThr	360
Db	288717	GAACACAGCTTCCGCGCGGCATCAGCCAGCGCAACTATCGGGCACTACTCGACA	288658
Qy	361	SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly	380
Db	288657	AGCAACGTTTCTTACAACCGTTCCACACCCAAACCTCGAAAAACCGCTTCTCCGCG	288598
Qy	381	GlyTyrThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu	400
Db	288597	GGCATCTGGTATGTGCGCGACCGCGCGGCATCGATCCAGGCTGGGGCGGAGTTTCTC	288538
Qy	401	AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet	420
Db	288537	GCAGAGCGCGGAAATCCCGCGCTCGGATATCGATTGGGCAACACGCCAGCCAGATG	288478
Qy	421	LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis	440
Db	288477	CTGACCGCTCTTGGAAACGCCAGCTCTCAACACCGTCTGCATCCGAAACGGCCAT	288418
Qy	441	TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle	460
Db	288417	TACTTCACGCGCAAAATCGGTACGACTTTTGGGCGCATTTCTGTCTCCACCGCGCTGATC	288358
Qy	461	ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe	480
Db	288357	CGCACCTCTGCCCGCGAGGTATTCTTCCACGCCGCAACCAAAACATCGGCAGCTTC	288298
Qy	481	IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly	500
Db	288297	ATCATACGCGACAAAGCGGTTACACCGTTCCCGCGCACAAATGCCCAACGTTCTCTCAGG	288238
Qy	501	LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly	520
Db	288237	CTGATGTTCCGACGCGCGCGCGTCTTCCGTGCGCGGTTACGAACCTCGACAGCATCGGG	288178
Qy	521	LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu	540
Db	288177	CTTGCGCGCGCAACGATCGGCTCTCCCGCAACGCGCCCTCTTGTGGCGACGCTGGAA	288118
Qy	541	TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla	560
Db	288117	TACCAACTCGCGTTTACGCGCACCTTTCGCGCGCGGTGTCCACGATATGGCGACGCC	288058
Qy	561	AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe	580
Db	288057	GCCCCCAATTTCAACAGTATGAAGCTGAACACGTTTCGGGACGTCGCGCTGGTTC	287998
Qy	581	SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg	600
Db	287997	AGCCCGCTCGCGCGTTTCTTCTCGACATCGCTACGGGCACGCGACAGAAATCCCG	287938
Qy	601	TrpHisIleSerLeuGlyThrArgPhe	609
Db	287937	TGGACATACGCTTGGGAACGCGCTTC	287911
LOCUS	AE002561	12429 bp	DNA linear BCT 25-MAY-2000

DEFINITION	Neisseria meningitidis serogroup B strain MC58 section 203 of 206 of the complete genome.
ACCESSION	AE002561 AE002098
VERSION	AE002561.1 GI:7227384
KEYWORDS	
SOURCE	Neisseria meningitidis MC58.
ORGANISM	Neisseria meningitidis MC58
REFERENCE	1 (bases 1 to 12429)
AUTHORS	Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C., Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F., Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D., Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D., Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E., Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H., Qin,H., Vanathevan,J., Gill,J., Scariato,V., Masignani,V., Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R., Rappuoli,R. and Venter,J.C.
TITLE	Complete genome sequence of Neisseria meningitidis serogroup B strain MC58
JOURNAL	Science 287 (5459), 1809-1815 (2000)
MEDLINE	20175755
PUBMED	10710307
REFERENCE	2 (bases 1 to 12429)
AUTHORS	Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C., Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F., Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D., Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D., Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E., Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H., Qin,H., Vanathevan,J., Gill,J., Scariato,V., Masignani,V., Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R., Rappuoli,R. and Venter,J.C.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..12429
	/organism="Neisseria meningitidis MC58"
	/strain="MC58"
	/db_xref="taxon:122586"
gene	/note="serogroup: B"
	187..1530
	/gene="NMB2129"
	187..1530
CDS	/gene="NMB2129"
	/note="similar to GB:L42023 SP:P44315 PID:1008797 PID:1205957 PID:121881 percent identity: 93.91; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="argininosuccinate synthase"
	/protein_id="AAF42437.1"
	/db_xref="GI:7227385"
	/translation="MSQHTILQSLPVQGVQVGFSGGLDTSAALLMKLGPVAY TANLGPDEDDYNAIPKAMETCAENARLIDCRAQLAHGEGIAAICQGFHVSFTGGIAY FNTPLGRAVDTMLVSKMEDDNIWNGDSTYKGNIDERYRGLLTPALKIYKPW LDOQFIDELGSRHEMSEFLIANGPNVKSVERAYSTDNNLGNATHEAKDLEFLNSGIK TVKPMGVAFWENVEVSEVSRPEEGVPVALNGKEYADVPVELFLEANRIGGRHGL GMSDQENRIIEAKSRGIYEPGMALFHAYERLVGTGHNEDTIEQYRINTYGRLLRLL YGRWFPDSQALMLRETAQRVAKAVTGEVTLLELRKNDYSILNTESPNITYPERLSM KEVEDAAFTPLDRIGQLTWRNLDITDTRVKLGITYSQSGLLSLGEGSVLPQLGNQK"
gene	1578..1820
	/gene="NMB2130"
	1578..1820
CDS	/gene="NMB2130"
	/note="hypothetical protein; identified by Glimmer2; putative"
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAF42438.1"

```
/db_xref="GI:7227386"
/translation="MIPVYKRNPMKEYKVIYQESLLSSLFFGAAKVNPINPSEPLNK
QTEPCWRVVTWEKOLRMLFFKREAYVVILEROV"
1813..2028
/gene="NBW2131"
CDS
1813..2028
/gene="NBW2131"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF42439.1"
/db_xref="GI:7227387"
/translation="MFKGVYTCGLGFAGWLLLVQLWFSLEALFKIITLWAGL
FVILIAALLVCGQYSEKKMKDDGIN"
complement(2104..3570)
/gene="NBW2132"
CDS
complement(2104..3570)
/gene="NBW2132"
complement(2104..3570)
/note="similar to PID:747922 percent identity: 48.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transferrin-binding protein-related protein"
/protein_id="AAF42440.1"
/db_xref="GI:7227388"
/translation="WFKRSVIAMACIFALSACGGGGSPDVKSADTLSKPAAPVSE
KETEAKEDAPQSGCGAPAGSQGSDMAVSEENTGNGAVTANAKNEDVEAONDM
PONAAGDSTTPHPYDPNMLANGNENQATDAGESQPAQPDMAAADMGQDDPSA
QGONAGTAAGCAQANNOAGSSDIPASNPAPANGSNFGRVDLANGVLIDGPSQ
NTTLTHCKGDSKGNFLDEEVLKSEFEKLSADAK ISNYKDKGDKRFVGLVADSVQ
MKGINQYIIFYKPKTSPARFRSRSRSLPAEMPLIPVQADTLIVDGEAVSLTGH
SGNIAPGPNRYLYGAEKLPGLSGSYALRVQCEPAKGEMLAGAAVNYEVLHFPTENG
RYPTRPFAARVDFGKSVSDGIIIDSDGLDMGTOKFAAIDGNGFKGTWTWENGSDV
SKGFYGAEEVAGYSYRPTDAEKGGFGVFAGKKEQD"
complement(3825..5054)
/gene="NBW2133"
CDS
complement(3825..5054)
/gene="NBW2133"
/note="similar to SP:P42602 PID:606028 GB:U00096
PID:1789473 percent identity: 82.74; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sodium/dicarboxylate symporter family protein"
/protein_id="AAF42441.1"
/db_xref="GI:7227389"
/translation="MAFGKSLFHAICRVSLVRQIAAGLALGIVIGSVSPQLGAAGLF
GSLFVGLKAVAPLVFVILVAATIAOHOKGNKAHRIPTIVLVIGTFSALTAVIAGM
VPTTHVLGAGDVAAPPSGIVEVLKSLMLNLANPINAANANYIGTLAWALVGA
ALRNHGSVDYTRQVADLAERAVSTVVKWIIRFAPLGFGLYSSTIAETGFGALAGAKL
LAVLGCMATFALANPAIVNWKIRNPPLVFTCLRESGVYAFFTRSAANIPVMA
LAKLGLHEDTYSISIPLAGTINMAGAAITITVLAAMAAHGTGIVDFATALLSLVA
TVSACGAGVAGGSLIPLACSLFGIDNVAMQVAVGVFIIGVQDSAEATLNSSTD
VFLTAADLGRGNRAE"
5527..7434
/gene="NBW2134"
CDS
5527..7434
/gene="NBW2134"
/note="conserved hypothetical protein"
Glimmer2: putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42442.1"
/db_xref="GI:7227390"
/translation="WPSEASRPVRTVLKSKAATPMHDTMTMKPTALLLPALFFPPH
AYAPADLSENKAGFALPKNSPTESVKLRKPPVLIDTQDSEIKDWHEHLPLIT
QQQEEVLDEKQFGLAEAPDNVKTMLRSKGFSSKSUTEKDGAYTHITPGPRKI
ANVWGLADLIDSLNLAERYNLENLMOQPVGSDFDQDSWENSKTSVLGAYTRKPY
LAKLGNTOAANVPDPTADLNVVDSDGRPIAFGDFEITGTQRYPEQIVSGLARFQPM
PYDLDLLDFQALEQNGHYSASVQADFRLQSDRPVVKVSVTEVKRHKLETGRLD
```

```
SEYLGCGKIADYNNLFNKGYIGSVVWMDMKYETTTLAAGISQPRNYRGNYTMSVNSY
KSTONLKRFPSSGGVYVDRAGIDARGLAEFRKIPGSAGVYDHSNATMLTAS
WRQLLNVLHPFENGHYLDGKTGTTGLTFLSSTALIRTSARAGYFETPNKKGITFI
RCOAGTYTARDNADVPISGLMFRSGGASSVRGYELSDISGLAGPNSVLPERRALLVGSLE
YOLPRTSLSGAVFHDMDGDAANFRKMLKLGSLGVRWFSPFLAPFSDIATYGHSDKK
IRWHISLGRF"
7441..11658
/gene="NBW2135"
CDS
7441..11658
/gene="NBW2135"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42443.1"
/db_xref="GI:7227391"
/translation="MATSDEIAANHFTDMDTDTPTDPTENGTNRKMPSEHRPTTP
AKRRRLKLSAALLSVLILAVFCFLWAGTEAGLFGLYQIPSMFVGNISQNLKGT
LDDGFDGONWSJETEGADLKISRFRFAMKPSLMRSLHITISAGDIAIVTKPTPK
EERPLSLPDSIDLPAAYVLDREFTGKISMGRKAFDKOTVYLERLDAASYRDKGRHLD
LKAADTPHSSSSGAASVGLKPFALDTALYTKGLEGKTIHSTARLSGLKDVRAELA
IDGNIIRLSKSVIHPFAESLDTLEEVLVKGFNINPAAFVPSLPDAGUNFDLTAIPS
FSDGIALGSDLENTKAGFADNRGIPVQVGLGVIRQDGTVHGTHTSAALLRGGI
RLSGKIDTEKIDILNIGINSVGAEDVPTAFKGLDGSIGGIGTASPKISWOLGIG
TARTDGSIAIASDPANGORKLVDTVNIAAGOSLTAQGYLELFDRLKLDIRSRFL
DPSRIDLDPOLPAGNINGSINLACELAKEFTGKMRPLPCTFNGVPITAGSADIYVESRAH
PRAAVDLRLGRNLIITDGGFGKGRDLNINITAPDLRSFGFLAGLSLNVGRHLSGDL
GGITFEDTDLGAARNLHIGKAADIRSLDFTLKGSPDTSRPIRADIKGRLSLSGGA
VIVTDLMLDGTGVQHRIRTHAAMTLDKPPFDLDASGGIINRELTWMKSGTIGLDIG
AELNKLKRPMPFHLNLEAGARVAASAMQAMGSLNLOHFESWDKTGTISAKGAAGHLHI
GAFNLKFPFPHNLVNGWDVAVYGRNAGYLNISROSGDAVLPGGALGLNGLSLK
TFQNDRTGILLDGGARFIRIADIGIANAFGCGNMANAPLGGRITASLIDIGALKPFL
PAAQNTIGSLNAAQIGRVGSPSVNAAVNGSNYNGKINITVCGSRSDFTPLGG
RNLNVADEAFRNFLPVGTVKSLNAAVTILGGSITADPHLGGSLNDRKLYRNQTGG
IILDNGSLRSHIAGRKWVDSLKFRHETGAEISVGSMENSGPDVDIGAVFPKRYLIS
RPNRLTVSGNTRLYRSPQKISVTGMIKTDOGLFGSKSSMPVSGDDVVVLGEVKE
AAAPLVNMLTDLNDGIRFAGYADVTIGKLTLTAGSGGSRVGRVTVIKGRYK
AYGODLDTKGTVSFVGLNDPNDNIRAPERLSPVGAEGEILGSLNSPRITLANEPM
SEKDLXSLILNRRAGSSGSDNAAALSAAGALLAGQINDRIGLVDLGGTFSKSRNAQ
TGEINPAEQVLTGKLTGKLYGYESISSAEQSVKLIYRLITRAITQAVARIGSRSSG
GELTYTTRDFRFSGDKDSAGNGK"
BASE COUNT 2986 a 3768 c 3144 g 2531 t
ORIGIN

Alignment Scores:
Pred. No.: 4,48e-201 Length: 12429
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 1 Gaps: 0

US-09-857-669-2 (1-609) x AE002561 (1-12429)
QY 1 MetWettLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
|||||
Db 5605 ATGATGATCAACCGACCGCCCTGCTCCTCCCGGCTTTATTTCTTCTCCGACGCATAC 5664
|||||
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
|||||
Db 5665 GCGCGCTGCCCGCACCTTTCGAAACAAACAGCGCGGGTTCGCATTGTTCAAAACAAA 5724
|||||
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValAlaGlyLeuGln 60
|||||
Db 5725 AGCCCCGACACCGAATCAGTCAAAATTAACCCCAAAATTCCTCCCTCATCGACACGAG 5784
|||||
QY 61 AspSerCluLysAspMetValGluGluHisLeuProLeuLeuThrGlnGlnGlu 80
|||||
Db 5785 GACAGTGAATCAAGATATGGTCGAAGAACACCTCGCGCTCATCACGACGACGAA 5844
|||||
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
|||||
```

```
Db 5845 GAAGTATTGGACAGGAAACAGACGGGCTTCCTCCGCCAAGAACGCCGCGCAACGTTAAA 5904
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 5905 ACGATGCTCCGCACCAAGGCTATTTCAGCAGCAAAAGCTAGCCCTGACGGAAAGACGGGA 5964
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 5965 GCTTATACGGGTACACATCACACCGGGCCCGCACCAAAATCGCCACGTCGGCGTCGCC 6024
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 6025 ATCTCGCGGACATCTCTTTCAGACGGCAAGCTTCGCCGAATACTACCGCAACGCGCTGAA 6084
QY 161 AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 6085 AACTGGCAGCAGCGGTAGGACGATTTTCGATCAGGACAGTTGGGAAACACGCAAAACT 6144
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 6145 TCGGTCTCGCGCGGTAAACGCGCAAAAGCTACCCGCTTGCCAAAGCTCGGCATACCCAG 6204
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 6205 CGGCGCGTCAACCCCGATACCGCCACCGCGGATTTGAACGTCGTCGTGGACAGCGCGCG 6284
QY 221 ProfileAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 6265 CCCATCGCTTCGGCGACTTTGAATATCCCGGCACACAGCGTTATCCCGCAAAATCGTC 6324
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 6325 TCGGCGCTTCGCGCTTCCAGCGCGGTATCGCGTACGACCTCGACCTGCTCGACTTC 6384
QY 261 GlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 6385 CAACAGCGCTCGAACAACAAAGCGGCATTTATTCGCGCGCGTCCGTCACAAAGCGGACTTCGAC 6444
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 6445 CGCTCCAAAGCGGACCGCGTCCCGTCAAAGTCAGCGTACCGAGGTCAACAGCCACAAA 6504
QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
Db 6505 CTCGAAACCGCATCCGCTCGATTCGGAATACGTTTGGCGCGCAAAATCGCCTACGAC 6564
QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 6565 TATTACAACTCTTCAACAAAGGCTATATCGGTTTCGCTCGGTATATGGACAAATAC 6624
QY 341 GluThrThrLeuAlaLagIleSerGlnProArgAsnTyrArgGlyAsnTyrTyrTyrThr 360
Db 6625 GAACACACGTTGCGCGCGGATCAGCCACCGCGGCAACTATCGGGCAACTACTTGACA 6684
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 6685 AGCAACGTTTCTTACACCGCTTCGACCAACCAAAACCTCGAAAAACCGCCTTCTCCGCG 6744
QY 381 GlyIleThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 6745 GCGCTGCGTATGTCGCGCAGCGCGCGGATCGATCCGAGCTGGGGGGGGAATTTCTC 6804
QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 6805 GCAGAACGCGGAAATATCCCGGCTCGGCTGCTCGATTGGGCAACACCCAGCCACCATG 6864
QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
Db 6865 CTGACCGCTCTTGGAAACGCCAGCTCTCAACAAACGCTGCTGATCCCGAAACGGCCAT 6924
QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 6925 TACTCGACGGCAAAATTCGGTACGACTTTGGGCACATTTCTGTCTCTCCACCGCGCTGATC 6984
```

```
QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 6985 CGCACTCTGCCGTCGAGGTTATTTCTTACGCCCGCAAAACAAAACCTCGGACGCTTC 7044
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaAlaArgAspAsnAlaAspValProSerGly 500
Db 7045 ATCATACGCGACAAAGCGGTTACACGTTGCCCGGACAATGCCGACGTTCTCTTCAGGG 7104
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 7105 CTGATGTTCCGCAAGCGCGCGCTTCTTCCGTCGCGGTTACGAACTCGACAGCATCGGA 7164
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 7165 CTGCGCGCGCGCAACGAGATCGGTCCTGCCGACGCCGCCCTCTCTGCGGAGCGCTCGAA 7224
QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 7225 TACCACTGCGGTTTACGCGCACGCTTTTCCGCGCGGTGTTCACGATATATGGGCGATGCC 7284
QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 7285 GCGCCCAATTTCAAACGATATCAAGCTGAAACACGCTTTCGGACTGGCGCTGCTGTTTC 7344
QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 7345 AGCCCGCTTGGCGCGCTTTTCTTCGACATCGCCTACGGGCACGCAATAGAAAATCCGC 7404
QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 7405 TGGCACAATCAGCTTGGGAACGCGCTTC 7431
RESULT 5
AX044035 AX044035 172325 bp DNA linear PAT 24-NOV-2000
LOCUS Sequence 114 from Patent WO0066791.
DEFINITION AX044035
ACCESSION AX044035
VERSION AX044035.1 GI:11342919
KEYWORDS Neisseria meningitidis.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
REFERENCE 1 (bases 1 to 172325)
AUTHORS Pizsa, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M.,
Scarlato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.
TITLE Neisseria genomic sequences and methods of their use
JOURNAL Patent: WO 0066791-A 114 09-NOV-2000;
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source Location/Qualifiers
1.172325
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/notes="sequence too long, cut in 8 pieces.-seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 349980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1349980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
2272325 172325 bases"
BASE COUNT 43072 a 47583 c 41465 g 40205 t
ORIGIN
Alignment Scores:
Pred. No.: 1,41e-199 Length: 172325
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 6 Gaps: 0
```

US-09-857-669-2 (1-609) x AX044035 (1-172325)

QY 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
|||||
Db 141935 ATGATGATCAAAACCGACCGCCCTGCTCTCGCGGCTTTATTTTCTTCGCGACGCATAC 141994

QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
|||||
Db 141995 GCGCCTGCGCGCGACCTTTCCGAAACAAAGCGCGGGTTTGCATTTGTTCAAAACAA 142054

QY 41 SerProAspThrGluSerValLysLysLeuLysProLysPheProValArgIleAspThrGln 60
|||||
Db 142055 ACCCGCGACCGAATCACTCAAAATTAACCCCAATTCCTCGCTCTCAGACAGCGAG 142114

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
|||||
Db 142115 GACAGTGAATCAAGATATGCTCGAAGAACACCTGCCGCTCATCAGCGAGCAGGAA 142174

QY 81 GluValLeuAspLysGluThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
|||||
Db 142175 GAAGTATTGGACAAGGAACAGCGGCTTCTCGCGAAGAGCGCGGACAACGTTAA 142234

QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrLysAspGly 120
|||||
Db 142235 AGATGCTCCGCGAGCAAGGCTATTTCAGCAAGAGTCAGGCTGACGGAAACACGGA 142294

QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
|||||
Db 142295 GCTTATACGGTACATACACCGGCGCGCCACCAAAATGCCAACGTCGCGCTCGCC 142354

QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
|||||
Db 142355 ATCCTCGCGGACATCTTTTCAGACGGCAACCTCGCGGAATATACCGCAACGCGCTGGA 142414

QY 161 AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
|||||
Db 142415 AACTGCGACGACCGCGTAGCGAGCATTTTCGATCAGGACAGTTGGGAAACACGCAAACT 142474

QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
|||||
Db 142475 TCCGTCTCGCGCGGTACCGCGCAAGGCTACCGCTTGCCAGCTCGGCAATACGCGAG 142534

QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
|||||
Db 142535 GCGGCGGTCAACCCCGGATACCGCCACCGCGCATTTGAAGTCGTCGTGACAGCGCGCGC 142594

QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
|||||
Db 142595 CCCATCGCCTTCGGCGGACTTTGAATTCACCGGCACACAGCGTTACCCCGCAACAAATCGTC 142654

QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
|||||
Db 142655 TCCGGCCTTCGGGTTTCCAGCCGGTATGCGGTAGGACCTCGACCTGCTGCTCGACTTC 142714

QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
|||||
Db 142715 CAACAGCGCTCGAACAACCGGCATTTTCGCGCGCTCGGTACAAACCGCGATTCGAC 142774

QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
|||||
Db 142775 CCGCTCAAGGCGACCGCTCCCGGTCAAAAGTCAGCGTCAACGAGGTCAACGCCCAAA 142834

QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
|||||
Db 142835 CTCGAACCGGCATCGCCCTCGAATTCGGAATACGGTTTGGCGGCAAAATCGCTACGAC 142894

QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
|||||
Db 142895 TATTACAACCTCTTCAACAAGGCTATATCGGTTCGGTTCGGTATGACAAATAC 142954

QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
|||||
Db 142955 GAACACACGCTTTCGCGCGGATCATGCCAGCGCGCAACTATCGGGCAACTACTGGACA 143014

QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
|||||
Db 143015 AGCAACGTTTCTACAAACGTTTCGACCCCAACCTCGAANAACGCGCTTCCTCCGGC 143074

QY 381 GlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
|||||
Db 143075 GCGCTGTGTATGTGCGCACCGCGGCGCATCGATGCCAGGCTGGGGCGGAATTCCTC 143134

QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyLysAsnSerHisAlaThrMet 420
|||||
Db 143135 GCAGAAGCGCGGAAATCCCGGCTCGGTGCGATTTGGGCAACAGCCACCGCATG 143194

QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
|||||
Db 143195 CTGACCGCTCTTGGAAACGCCAGCTGCTCAACAGCTGCTGCATCCCGAAACGCCAT 143254

QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
|||||
Db 143255 TACCTCGACGCGCAAAATCGTACGACTTTGGGCACATTCCTGCTCCACCGCGCTGATC 143314

QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
|||||
Db 143315 CGCACTCTGCGCGGTGACGTTATTCTTCACGCGCGAACAACAACTCGGCACGTTTC 143374

QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
|||||
Db 143375 ATCATACGCGGACAGCGGTTACACCGTTGCCGCGACANTGCCGACGTTCTCTCAGGG 143434

QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
|||||
Db 143435 CTGATGTTCCGACGCGCGCGCTTCTCGTCCGCGGTTACGAACTCGACAGCATCGGA 143494

QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
|||||
Db 143495 CTTCGCGCGCGCAACGATCGGTCTGCGCGAACGCGCTCTCTGTGGCGACCTGGAA 143554

QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
|||||
Db 143555 TACCAACTCGCGTTTACGCGCACCTTTCCGCGCGGTGTTCCACGATATGGCGATGCC 143614

QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
|||||
Db 143615 GCGGCAATTTCAAACGATGAAGCTGAAACACGTTTCGGAGCTGGCGCTGCGCTGGTTC 143674

QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
|||||
Db 143675 AGCCGCTTTCGCGCGTTTCTTCGACATCGCTACGGGCACAGCGATAAGAAATCCGC 143734

QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
|||||
Db 143735 TGGCACATCAGCTTGGGAACGCGCTTC 143761

RESULT 6

LOCUS AX026692 1764 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 5 from Patent WO0034480.
ACCESSION AX026692
VERSION AX026692.1 GI:10187862
KEYWORDS
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 1764)
AUTHORS Ruelle,J.L.
TITLE Polypeptides and polynucleotides basb040 from neisseria
meningitidis and vaccine comprising said polypeptides and
polynucleotides
JOURNAL Patent: WO 0034480-A 5 15-JUN-2000;
FEATURES Location/Qualifiers
1. .1764

REFERENCE AUTHORS	Ralstonia. 1 (bases 1 to 189050) Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L., Chandler,M., Choisine,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Siquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.
TITLE JOURNAL	Genome sequence of the plant pathogen Ralstonia solanacearum Unpublished
REFERENCE AUTHORS	2 (bases 1 to 189050) Boucher,C.A.
TITLE JOURNAL	Direct Submission Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT	Christian.Boucher@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES source	Location/Qualifiers 1..189050 /organism="Ralstonia solanacearum" /strain="GMI1000" /db_xref="taxon:305" complement(101..982) /gene="RSC2228" /note="RSC1371" complement(101..982) /gene="RSC2228" /EC_number="3.1.1.45" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Homology predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PUTATIVE CARBOXYMETHYLENEBUTENOLIDASE (DIENELACTONE HYDROLASE). PROTEIN" /protein_id="CAD15935.1" /db_xref="GI:17429248"
gene	/translation="MHLEDAEVSFLVPSFRSFRSFKTALGSFAAAVLPSQAIIH TDFNGLTGVETVPVPGVKVPAYRAQPEKGTGLPMVTWSEIFGVHEYIADVCRRRAK LGYLITFCWCGRQAWLFAEHSFDIKAAVWVYHLKGNATALQAFPIDKVADLKAPVL GLYGAKDTGITQDAVQAAMKALEASANPRAKASRLIIVPDAGHAFHADVNPSTYKADA EDGWKKCLAWFKDHGVA" /gene="ndh" /note="RSC2229; RSC1370" complement(1026..2351) /gene="ndh" /EC_number="1.6.99.3" /function="small molecule metabolism; energy metabolism, carbon: aerobic respiration" /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
gene	/function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
CDS	/translation="MDSNVVTGTPDGVSDNGASAEAGGVAKQVAGVSAEELIGHQPAHQ TTADMDVNTVGAEHRADAAKSDAAHRAERKMIPSLRALADEARATKARASWQ MFTIMAEFIETAEYLSIRPAVSITYGSARLSESPYQRTIDIAFLFSAGFAVISGG GPGIMEAANKGAHGKKSASGLNIELPHEOQGNPYODISMRFFFTFRKVTFTVKNSDA FIVMGGCGTLELAEVLTLVOTCKSRSPVPMVFGSHFWKGLLDWFRFTLLPMGLIAE HDLDMIRIVDEPKDALDAVTEFEKREGVSPIPKKEMFYL" 6411..6830 /gene="RSC2232" /note="RSC1367"
gene	/db_xref="GI:17429249" /translation="MAGOVDCAGGPRIVVVGAGGAGLELATRLGDKLAKRAARVYL DRNPTHWKLLHEVAQSDMDPNTHOLEYAAORWGHFEFOGELKGLDRAAKTITVS GCVDDAGTEVLPERATAYDMLVLAIGSVTHFGVPGCTAEHAJALDAASQAEFRERKLI SACRAONGVGDARAQVDIAIVGAGATGVLSNLAELNTHAVLAAYGLJHKLDPRLDIRI HLIEGSPRIILASERVSTETKLLQKLNVEVITGERVNOVTDASVAKTASGKSIIPADL TVMAAGIRAPATILGELGLPVNKLGVVVSRTLQAEGDDAIYAFGDCASCPWPEASAY PPRAOAAHQOATLYKALCRLECKPVEPFGFKDLGSLVSLGHFSAVGSMLGMLGGTIGGT MFIEGMARLWYTSLYRMVMAHLHGFVRMALDVTVTHWLSRKTNPVKLH" complement(2536..5376) /gene="poia" /note="RSC2230; RSC1369" complement(2536..5376) /gene="poia" /EC_number="2.7.7.7" /function="macromolecule metabolism; macromolecule synthesis, modification; dna - replication, repair, restr./modif." /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PROBABLE DNA POLYMERASE I PROTEIN" /protein_id="CAD15937.1" /db_xref="GI:17429250" /translation="MGMSQETLLLVGSSYLRYAYHALPDLRNGEGFPPTGAIYGI NMLRLNDYPAKYSACVFDKAGKTRFDELYPAYKEHRAPMDLRLQOETPEHRAVRA LWGPILMVDGVEADDVIGTIAEARAAREGIFTVVSTGDKDLAOLVNDHVTLVNTMSNT LDPAGVLAKEGVPCKIYDLSLIGDVTNVPCVPKGTAKVKLGEVGSGLDSVIAR AGEIKVGVNGENRLNTDMLPKGRELLTVTKDCULATEVPTFEALVDGDEGRKLVDFP GRYGEKTLREAGSESLPDRSVGAARKAATPVKYAPEAAQPOAGLDFEAPAEV KYETVTTAEQLSEMMRLIDASELVICIDITETTSIDPMLAQGLVLSVEPQACYIPVA HRGPDVSDAAGOLSRCAVLMKPMLEDAARKKQGLKYLDAHFVANHGLARGLIA HDTMLQSVLASHRNHGMALAEKRLVHLKTIITYEEVCGKASQIGEPDEVLDRATEYA AEDADITLRLHRLYPKVAEDDKLRYVEQIEMPTSIYLOKMERNGVLDIAERLARQS TGLGRLLEEQAYALAGAPFNLSPKQIGEIFFTOLKLPVTKKTAGAPSTDEELV OKLAEDYPLPKLLDLDFGLAKLSTYDKLPKMNVPQTGRVHTTYGOATVATGRLAST DPLNQNPVTRTEEGRRIRAEVAPAGSVISADYSQTELRINAHLSDEGLLRAREG QDVHRTAAAEVFGVTPLEVTAEQRRAKVTINFGLIYGMSAFGLSNLIGRDAAKLYI DHYFARYGVARYNDETRLVAHERGVYETVGRRLPLDINGNGGPRROAASRAALIA PMQGTADLLKLSMLAVQDWLDGSPSAGKPGAPMTRQVMQVHDELVLVLEPQBELSLV RERLPELMCGVASURVPLVAEVGVGVANWEEAH" 5375..6328 /gene="RSC2231" /note="RSC1368" 5375..6328 /gene="RSC2231" /function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
CDS	/translation="MDSNVVTGTPDGVSDNGASAEAGGVAKQVAGVSAEELIGHQPAHQ TTADMDVNTVGAEHRADAAKSDAAHRAERKMIPSLRALADEARATKARASWQ MFTIMAEFIETAEYLSIRPAVSITYGSARLSESPYQRTIDIAFLFSAGFAVISGG GPGIMEAANKGAHGKKSASGLNIELPHEOQGNPYODISMRFFFTFRKVTFTVKNSDA FIVMGGCGTLELAEVLTLVOTCKSRSPVPMVFGSHFWKGLLDWFRFTLLPMGLIAE HDLDMIRIVDEPKDALDAVTEFEKREGVSPIPKKEMFYL" 6411..6830 /gene="RSC2232" /note="RSC1367"
gene	/db_xref="GI:17429249" /translation="MAGOVDCAGGPRIVVVGAGGAGLELATRLGDKLAKRAARVYL DRNPTHWKLLHEVAQSDMDPNTHOLEYAAORWGHFEFOGELKGLDRAAKTITVS GCVDDAGTEVLPERATAYDMLVLAIGSVTHFGVPGCTAEHAJALDAASQAEFRERKLI SACRAONGVGDARAQVDIAIVGAGATGVLSNLAELNTHAVLAAYGLJHKLDPRLDIRI HLIEGSPRIILASERVSTETKLLQKLNVEVITGERVNOVTDASVAKTASGKSIIPADL TVMAAGIRAPATILGELGLPVNKLGVVVSRTLQAEGDDAIYAFGDCASCPWPEASAY PPRAOAAHQOATLYKALCRLECKPVEPFGFKDLGSLVSLGHFSAVGSMLGMLGGTIGGT MFIEGMARLWYTSLYRMVMAHLHGFVRMALDVTVTHWLSRKTNPVKLH" complement(2536..5376) /gene="poia" /note="RSC2230; RSC1369" complement(2536..5376) /gene="poia" /EC_number="2.7.7.7" /function="macromolecule metabolism; macromolecule synthesis, modification; dna - replication, repair, restr./modif." /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PROBABLE DNA POLYMERASE I PROTEIN" /protein_id="CAD15937.1" /db_xref="GI:17429250" /translation="MGMSQETLLLVGSSYLRYAYHALPDLRNGEGFPPTGAIYGI NMLRLNDYPAKYSACVFDKAGKTRFDELYPAYKEHRAPMDLRLQOETPEHRAVRA LWGPILMVDGVEADDVIGTIAEARAAREGIFTVVSTGDKDLAOLVNDHVTLVNTMSNT LDPAGVLAKEGVPCKIYDLSLIGDVTNVPCVPKGTAKVKLGEVGSGLDSVIAR AGEIKVGVNGENRLNTDMLPKGRELLTVTKDCULATEVPTFEALVDGDEGRKLVDFP GRYGEKTLREAGSESLPDRSVGAARKAATPVKYAPEAAQPOAGLDFEAPAEV KYETVTTAEQLSEMMRLIDASELVICIDITETTSIDPMLAQGLVLSVEPQACYIPVA HRGPDVSDAAGOLSRCAVLMKPMLEDAARKKQGLKYLDAHFVANHGLARGLIA HDTMLQSVLASHRNHGMALAEKRLVHLKTIITYEEVCGKASQIGEPDEVLDRATEYA AEDADITLRLHRLYPKVAEDDKLRYVEQIEMPTSIYLOKMERNGVLDIAERLARQS TGLGRLLEEQAYALAGAPFNLSPKQIGEIFFTOLKLPVTKKTAGAPSTDEELV OKLAEDYPLPKLLDLDFGLAKLSTYDKLPKMNVPQTGRVHTTYGOATVATGRLAST DPLNQNPVTRTEEGRRIRAEVAPAGSVISADYSQTELRINAHLSDEGLLRAREG QDVHRTAAAEVFGVTPLEVTAEQRRAKVTINFGLIYGMSAFGLSNLIGRDAAKLYI DHYFARYGVARYNDETRLVAHERGVYETVGRRLPLDINGNGGPRROAASRAALIA

REFERENCE AUTHORS	Ralstonia. 1 (bases 1 to 189050) Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L., Chandler,M., Choisine,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Siquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.
TITLE JOURNAL	Genome sequence of the plant pathogen Ralstonia solanacearum Unpublished
REFERENCE AUTHORS	2 (bases 1 to 189050) Boucher,C.A.
TITLE JOURNAL	Direct Submission Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT	Christian.Boucher@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES source	Location/Qualifiers 1..189050 /organism="Ralstonia solanacearum" /strain="GMI1000" /db_xref="taxon:305" complement(101..982) /gene="RSC2228" /note="RSC1371" complement(101..982) /gene="RSC2228" /EC_number="3.1.1.45" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Homology predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PUTATIVE CARBOXYMETHYLENEBUTENOLIDASE (DIENELACTONE HYDROLASE). PROTEIN" /protein_id="CAD15935.1" /db_xref="GI:17429248"
gene	/translation="MHLEDAEVESLVPSRFSRFSFKTALGSFAAAVLPSQAIIH TDFNGLTGVETVPVPGVKVPAYRAQPEGKGTGLPMVTWSEIFGVHEYTADVCRREAK LGYLITFCWCGRQAWLFAEHSFDPDKAAVWVYHLKGNATALQAFPIDKVADLKAPVL GLYGAKDTGITQDAVQAAMKALEASANPRAKASRLIIVPDAGHAFHADVNPSTYKADA EDGWKKCLAWFKDHGVA" /gene="ndh" /note="RSC2229; RSC1370" complement(1026..2351) /gene="ndh" /EC_number="1.6.99.3" /function="small molecule metabolism; energy metabolism, carbon: aerobic respiration" /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
gene	/function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
CDS	/translation="MDSNVVTGTPDGVSDNGASAEAGGVAKQVAGVSAEELIGHQPAHQH TTADMDVNTVGAEHRADAAKSDAAHRAERKMIPSLRALADEERATAKARASWQ MFTIMAEFTEATEYLSITRPASVITYSGARLSESPYQRTIDTARLFSDAGFAVTSGG GPGIMEAANKGAHGKKSASGLNLTLPHEQOQNPYODISMRFFFTFRKVTFTVFKNSDA FIVMGGCGTLDLAEBVLTIVOTCKSRSPVPVWFGSHFWKGLLDWFRFTLLPMGLIAE HDLDMIRIVDEPKDADNAVTEFEKREGVSPIPKKEMFYL" 6411..6830 /gene="RSC2232" /note="RSC1367"
gene	/db_xref="GI:17429249" /translation="MAGOVDCAGGPRIVVVGAGGAGLELATRLGDKLRRAARVYLIV DRNPTHWKLLHEVAQSDMDPNTHOLEYAAORWGHFEFOGELKGLDRAAKTITVS GCVDDAGTEVLPERATAYDMLVLAIGSVTHFGVPGCTAEHAJALDAAQSOAERFERKLI SACRAONGVGDARAQVDIAIVGAGATGVLSNLAELRNTAHVLAAYGLJHKLDPRLDIRI HLIEGSPRIILASERVSTETKLLQKLNVEVITGERVNOVTDVAVKTKASGKSIIPADLI TVMAAGIRAPATILGELGLPVNKLGVVVSRTLQAEGDDAIYAFGDCASCPWPEASAY PPRAOAAHQOATLYKALCRLECKPVEPFGFKDLGSLVSLGHFSAVGSMLGMLGGTIGGT MFIEGMARLWYTSLYRMHVMLHGFVRMALDVTVTHWLSRKTNPVKLH" complement(2536..5376) /gene="poia" /note="RSC2230; RSC1369" complement(2536..5376) /gene="poia" /EC_number="2.7.7.7" /function="macromolecule metabolism; macromolecule synthesis, modification; dna - replication, repair, restr./modif." /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PROBABLE DNA POLYMERASE I PROTEIN" /protein_id="CAD15937.1" /db_xref="GI:17429250" /translation="MGMSQETLLLVGSSYLRYAYHALPDLRNGEGFPPTGAIYGI NMLRLNDYPAKYSACVFDKAGKTRFDELYPAYKEHRAPMDPLRQOETPEHRAVRA LWGPILMVDGVEADDVIGTIAEARAAREGIFTVVSTGDKDLAOLVNDHVTLVNTMSNT LDPAGVLAKEGVPCKIYDLSLIGDVTNVPCVPKGTAKVKLGEVGSGLDSVIAR AGEIKVGVNGENRLNTDMLPKGRELLTVTKDCULATEVPTFEALVDGDEGRKLVDFP GRYGEKTLREAGSESLPDRSVGAARKAATPVKYAPEAAQPOAGLDFEAPAEV KYETVTTAEQLSEMMRLDASLVICIDTETTSIDPMLAQGLVLSVEPQACYIPVA HRGPDVSDAAGOLSRCAVLMKPMLEDAARKKQHLKYDAHFVANHGLRGLIA HDTMLOSYVLASHRNHGMALAEVRLHKTITYEEVCGKASQIGEPDEVLDRATEYA AEDADITLRLHRLYPKVAEDDKLRYVEQIEMPTSIYLOKMERNGVLDIAERLARQS TGLGRLLEEQAYALAGAPFNLSPKQIGEIFFTOLKLPVTKKTAGAPSTDEELV OKLAEDYPLPKLLDLPFLGAKLSTYDKLPKMNVPQTRVHTTYGOATAVTGRLAST DPLNQNPVTRTEEGRRIRAEVAPAGSVISADYSQTELRIMAHLSDEGLLRAFREG QDVHRATAEVFGVTPLEVTAEQRRAKVFNGFLIYGMSAFGLSNLGRDAAKLIY DYFRARYGVARYNDETRLVAHERGVYVETGRLMLPDINGNGGPRROAASRAALIN PMQGTADLLKLSMLAVQDWLDGSPSAGKPGAPMTRQVWQVHDELVLVLEPQBELSLV RERLPELMCGVASURVPLVAEVGVGVANWEEAH" 5375..6328 /gene="RSC2231" /note="RSC1368" 5375..6328 /gene="RSC2231" /function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOHETICAL PROTEIN" /protein_id="CAD15938.1" /db_xref="GI:17429251"
CDS	/translation="MDSNVVTGTPDGVSDNGASAEAGGVAKQVAGVSAEELIGHQPAHQH TTADMDVNTVGAEHRADAAKSDAAHRAERKMIPSLRALADEERATAKARASWQ MFTIMAEFTEATEYLSITRPASVITYSGARLSESPYQRTIDTARLFSDAGFAVTSGG GPGIMEAANKGAHGKKSASGLNLTLPHEQOQNPYODISMRFFFTFRKVTFTVFKNSDA FIVMGGCGTLDLAEBVLTIVOTCKSRSPVPVWFGSHFWKGLLDWFRFTLLPMGLIAE HDLDMIRIVDEPKDADNAVTEFEKREGVSPIPKKEMFYL" 6411..6830 /gene="RSC2232" /note="RSC1367"
gene	/db_xref="GI:17429249" /translation="MAGOVDCAGGPRIVVVGAGGAGLELATRLGDKLRRAARVYLIV DRNPTHWKLLHEVAQSDMDPNTHOLEYAAORWGHFEFOGELKGLDRAAKTITVS GCVDDAGTEVLPERATAYDMLVLAIGSVTHFGVPGCTAEHAJALDAAQSOAERFERKLI SACRAONGVGDARAQVDIAIVGAGATGVLSNLAELRNTAHVLAAYGLJHKLDPRLDIRI HLIEGSPRIILASERVSTETKLLQKLNVEVITGERVNOVTDVAVKTKASGKSIIPADLI TVMAAGIRAPATILGELGLPVNKLGVVVSRTLQAEGDDAIYAFGDCASCPWPEASAY PPRAOAAHQOATLYKALCRLECKPVEPFGFKDLGSLVSLGHFSAVGSMLGMLGGTIGGT MFIEGMARLWYTSLYRMHVMLHGFVRMALDVTVTHWLSRKTNPVKLH" complement(2536..5376) /gene="poia" /note="RSC2230; RSC1369" complement(2536..5376) /gene="poia" /EC_number="2.7.7.7" /function="macromolecule metabolism; macromolecule synthesis, modification; dna - replication, repair, restr./modif." /note="Product confidence : probable Gene name confidence : probable predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PROBABLE DNA POLYMERASE I PROTEIN" /protein_id="CAD15937.1" /db_xref="GI:17429250" /translation="MGMSQETLLLVGSSYLRYAYHALPDLRNGEGFPPTGAIYGI NMLRLNDYPAKYSACVFDKAGKTRFDELYPAYKEHRAPMDPLRQOETPEHRAVRA LWGPILMVDGVEADDVIGTIAEARAAREGIFTVVSTGDKDLAOLVNDHVTLVNTMSNT LDPAGVLAKEGVPCKIYDLSLIGDVTNVPCVPKGTAKVKLGEVGSGLDSVIAR AGEIKVGVNGENRLNTDMLPKGRELLTVTKDCULATEVPTFEALVDGDEGRKLVDFP GRYGEKTLREAGSESLPDRSVGAARKAATPVKYAPEAAQPOAGLDFEAPAEV KYETVTTAEQLSEMMRLDASLVICIDTETTSIDPMLAQGLVLSVEPQACYIPVA HRGPDVSDAAGOLSRCAVLMKPMLEDAARKKQHLKYDAHFVANHGLRGLIA HDTMLOSYVLASHRNHGMALAEVRLHKTITYEEVCGKASQIGEPDEVLDRATEYA AEDADITLRLHRLYPKVAEDDKLRYVEQIEMPTSIYLOKMERNGVLDIAERLARQS TGLGRLLEEQAYALAGAPFNLSPKQIGEIFFTOLKLPVTKKTAGAPSTDEELV OKLAEDYPLPKLLDLPFLGAKLSTYDKLPKMNVPQTRVHTTYGOATAVTGRLAST DPLNQNPVTRTEEGRRIRAEVAPAGSVISADYSQTELRIMAHLSDEGLLRAFREG QDVHRATAEVFGVTPLEVTAEQRRAKVFNGFLIYGMSAFGLSNLGRDAAKLIY DYFRARYGVARYNDETRLVAHERGVYVETGRLMLPDINGNGGPR

QY 362 AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly 381
Db 5224 GACCTGGTCGACACGAGAGCAAGCTGTCACCTT-----GGCGGC 5183

QY 382 IleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAla 401
Db 5182 GAATGGACACAGCAGCGCCGATGGTGGCAGCGGTGGTTCGCTGAACCTGGATCGC 5123

QY 402 GluLysArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu 421
Db 5122 GAGAGTACAGCTGGCGGAC-----GACTCCGGGTGAGCAGCTTCCTGATCGCG 5072

QY 422 ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyr 441
Db 5071 GGGATCGCTACTCGCTGCTGGAGCCGACACAAAGCTCGACCCACCATGGCTACCGC 5012

QY 442 LeuAspGlyLysIle----GlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu--- 459
Db 5011 CTGAGTTCACAGCTCAAGGGGGGGAAGAGCTGCTGGCGGACCGAGCTCCTCCAT 4952

QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
Db 4951 GTCGACGCCATGGCCAGGCGCTGACAGCTTCGCCGGCGCCATCGCTCTCGCG--- 4895

QY 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSer 499
Db 4894 -----CGCTCGAGTAGCGGGAATGCCACCAACACTACAGTGCATGCCGCC 4844

QY 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 4843 TCGCTGGCTTCTCGCTGGCGGACAGAGCTGGCTGGCTACGACTACCGGACGCTG 4784

QY 520 GlyLeuGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeu 539
Db 4783 TCGCCGGAGATTCGAGTGGCGACAGATCGCGCGCGCTACATGATCGCGCGAGGTC 4724

QY 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 559
Db 4723 GAGTATCAATATCGCTGGCGGACGCTGGCGCTGGCGCTGCTCGACCGGGAAC 4664

QY 560 AlaAlaAlaAsnPhelYsArgMetLysLysHisGlySerGlyLeuGlyValArgTrp 579
Db 4663 GCCCTCAACTCGCTGGACTTCCCGCTCATAGACCGGGGCTGGCTGGCGCTGG 4604

QY 580 PheSerProLeuAlaProPheSerPheAspIleAlaTyrGly---HisSerAspLysLys 598
Db 4603 GTCGCGCGGTGCGCCCTTTCGCTCGACCTGGCCCATCGCTCGACGACGCGCGGT 4544

QY 599 IleArgTrpHisIleSerLeuGly 606
Db 4543 TTCGCGCTGCACTTCTCCATGGGG 4520

RESULT 9
AX078576 1374 bp DNA linear PAT 22-FEB-2001
LOCUS
DEFINITION Sequence 90 from Patent WO0107624.
ACCESSION AX078576
VERSION AX078576.1 GI:13158218
KEYWORDS
SOURCE Pseudomonas putida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1374)
Fraser, C.M., Venter, C., Tuemmler, B., Hoheisel, J., Duesterhoeft, A., Hilbert, H., Timmis, K.N., Moore, E., Straetz, M., Heim, S., Nelson, K.E., Hickey, E. and Peterson, J.
Dna sequences which are suited for specifically detecting Pseudomonas putida K2440
Patent: WO 0107624-A 90 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US) ; QIAGEN GmbH (DE) ;
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;

Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische Hochschule Hannover (DE)
FEATURES
source 1. .1374
Location/Qualifiers
/organism="Pseudomonas putida"
/db_xref="taxon:303"
BASE COUNT 264 a 428 c 429 g 251 t 2 others
ORIGIN
Alignment Scores: 3.41e-17 Length: 1374
Pred. No.: 384.50 Matches: 115
Score: 42.52% Conservative: 81
Best Similarity: 24.95% Mismatches: 224
Query Match: 12.12% Indels: 41
DB: 6 Caps: 11

US-09-857-669-2 (1-609) x AX078576 (1-1374)

QY 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaVal 186
Db 43 GTGAACAACCTCAATCAGCGCATATTACAGATGCAAGCGGTGATCCAGAACCGCG 102

QY 187 ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
Db 103 TCGCGCTATGCTTCTTTCAGTGGCGCTTCAGTAGCCAGCGCTGGCGCTGCAGCCGCA 162

QY 207 ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAsp 226
Db 163 GCCGGTGGCGGATATCGAACTGGTCTACAGAGTGGCGCGCTTATCGCGCTGGGGCC 222

QY 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaPhe 246
Db 223 GTCACCTTCGTTGGCGACACCGCTCGACGAGACCTGCTGCAGCGCATGGTCTCGTTC 282

QY 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 283 AAGCGGGTACCCCTTACGACTCGGAATGTCGACAGCTGAACAACGACCTGCAATCG 342

QY 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArg 286
Db 343 AGCGGCTATTCGAAGCGTGGCGTGGATGCGCGCGCCACTGCTGTGGGCGAAGAA 402

QY 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyLea 306
Db 403 GTCCCGTGGATGTTATCTGGAACCCGTAAACCGACCGACCATGGCGCTTGGCGCTGGC 462

QY 307 LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn 326
Db 463 TTCCTCGACGAGCTCGGCGCGCGCAAGCCAACTGGACCGCCACTGGGTCAACCCA 522

QY 327 LysGlyTyrIleGlySerValValTrpAspMetAspLys----- 339
Db 523 CAAGGCCAC-----AGCTATGGCTGGGAAACCGAACTGTGGCGCCCGCGCAGAACGTC 576

QY 340 -----TyrGluThrThrLeuAlaAlaGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn 356
Db 577 GGCTGTGGTATGACATTCCTCGACCGCGCTGACCGCAAGCTGCCAGAGTGGCGCTTGGCGCGC 636

QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 637 GGCTACCAAGACGAGGAGCTT-----GCCGCGACCGACAGCTCGCAAGCTG 684

QY 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 685 TTGACGGTGGCGCCCGAGTGGCACAGCAAGCTGCCAGAGTGGCTGGCAGCGGTGATTCG 744

QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 745 CTCAAGTACCACGCGAAGAAATATCGCGCTG---GGTGCAGGACTCCGGTTTGGAGCAACCTG 801

QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436


```

CDS
/gene="ECs5010"
2954..3622
/gene="ECs5010"
/notes="similar to YJBE_ECOLI gi11790459 percent identity
96 in 222 aa (Conserved in E.coli K-12)."
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38433.1"
/db_xref="GI:13364487"
/translation="MVILPLMRRVVRKRPAPILICLLQACSAATTRELGNLSMDSLEFGT
PGVQLDDDDIQQNPYASQYMLNGPQLFVFLAFADGQKWKVTDQOATLVTHQGLRV
KTLIGDNLIEYNNLAADPLIKPAQIIDGATWMTMGWTEYQQVRYATARSFKWDGT
GTVKVGSDETAIVRLDEEVSTDAQRHNRVWIDSGGQIRQSEYLGADYFPVKTTIK
AAQ"
3619..4356
/gene="ECs5011"
3619..4356
/gene="ECs5011"
/notes="similar to YJBG_ECOLI gi11790460 percent identity
97 in 245 aa (Conserved in E.coli K-12)."
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38434.1"
/db_xref="GI:13364488"
/translation="MIKQTRIALVLLSVGSSVFAAGTVKVFNSGSEAKTLTGAERHLI
DLVQPLRANSMPGAVISEELATAALRQOQALLRLAQGGADSDAAALINLRQI
QIQLKVRQKINLDPDIVRAERANPLQNTYTLWVGPPPTVFLGLISRPQNPQ
FTPGROVASVLSQSLGGADRSAWVYPDGRTPKAPVAYWNRKRVPEMPGSIIVG
LADVNSEMPDALNADILQTLTORIQ"
4356..6452
/gene="ECs5012"
4356..6452
/gene="ECs5012"
/notes="similar to YJBH_ECOLI gi11790461 percent identity
99 in 698 aa (Conserved in E.coli K-12)."
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38435.1"
/db_xref="GI:13364489"
/translation="MKRRHLLSLLALIGSTACGYETPPIGPSQSDFGVGLLQTP"
ARWARGESLNRDNDQDRIYYSASVOLFPWLETTIRYTDVTRQYSSVEAFSGDQTY
KDAFDLKLRLWEESYWLQVAVGARDIGGTGLFDAEYLYVASKAWGPFDFTLGHWGY
LGTSGVKNPLCSADSKYCYRDNYSYQAGSIDGQSFHGPASLFGVEYOTPMOPLRL
KLEVEGNYOODPAGKLEOKSKFNVGAIYRTDADVNLSYERANTFMFGVTIURTNP
DLRPSYNDNARPOYQPODAILQHSVAVANOLTLKYNAGLADPQIOAKGDTLYVTGE
QVKYRSREGIVRANRIVMNDLFDGIKTIRVTENRNLMPQVTTETDVSLSKHLAGEP
LGHETLAKRRVEPVVPKSTEQGWYIDKSRFDHFDPLNLSVGGPENFYMQLYGVNG
TADMLTLHLTTGSLFANLANNYDKFNTNPPQDSHLPRVTHVREYVQNDYVNNL
QANYFOHLGDFGVQYGVGYLETMFGAGAEVLYRLPDSNMAFGLDANYKRDWMSRA
KDMKFTDYSVKTGHTAYWTPSPFADVLKVASVGYLAGDKGDTLEIAKRFDSGVVV
GGYATLNNYSKEEYGEEDFTKGYVSVPLDLFSSGSTRSAAIGWTPLTDGGQQQLGR
KPOLYDMTSDRSVNR"
7047..7457
/gene="ECs5013"
7047..7457
/gene="ECs5013"
/notes="similar to YJBA_ECOLI gi11790462 percent identity
98 in 136 aa (Conserved in E.coli K-12)."
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38436.1"
/db_xref="GI:13364490"
/translation="MTSLSRPRVEFISTILQTVNLGLCLGLILVYVFLGKGTVHLAD
VLFAPQASKEYELVEGLVYVLYFEFIALIVKYFQSGFHPRLRYFVYIGITAVRLII
VDHKSPLDVLIIYSAAILLVVTLWLCNKKLRKRE"

```

```

gene
complement(7501..8976)
/gene="ECs5014"
CDS
complement(7501..8976)
/gene="ECs5014"
/notes="similar to XYLE_ECOLI gi11790463 percent identity
100 in 491 aa (Conserved in E.coli K-12)."
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="xylose-proton symport"
/protein_id="BAB38437.1"
/db_xref="GI:13364491"
/translation="MNTQYNSSYIFSTLIVATLGLLFGYDTAVISCTVESLTVFVA
PONSLSAASNLGFCVASALIGCIIGGALGVCSNRFGRDRSLKTAAYLFTTSGVGS
ANPELGFTSINPDNTVPVLGVYVPEFVYRIIIGGIVGLASMLSPMYLAELAPAHIR
GKLYSPNQAIIFGQLLYVCVNYFIARSDASHLNDGHRYPFASCIIPALLFLMLLY
TVPSRWLMSRGKEQAEILKRKIMGNLTATQAOVEIKHSLDHGRKTRILLMFGVG
VIVIGYMLSIFOQFVGINVLYYAFVFTLGASTDIALIQTIIIVGINLTFTVLAIM
TVDKFRKPLQIITIGALGMAIGMFSLGTAFTYQAPGVALLSMLFYVAAFAMSGPVCW
VLLSEIFPNAIRKALAIATAAQAOLANYFVSWTFPMMDKNSLVVAHFHNGFSYWIYCC
MGVLAALFNKKVPETKGTLELEALWEPETKKTKQTATL"
complement(9348..10238)
/gene="ECs5015"
CDS
complement(9348..10238)
/gene="ECs5015"

Alignment Scores:
Pred. No.:      5.65e-14      Length:      243184
Score:          380.50      Matches:    145
Percent Similarity: 39.46%      Conservative: 87
Best Local Similarity: 24.66%      Mismatches:  245
Query Match:     11.99%      Indels:      111
DB:              1          Gaps:          20

US-09-857-669-2 (1-609) x AP002568 (1-243184)
Qy 74  LeuileThrGlnGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu 93
Db 211436 CTTTCTACGATTGAAGTGATGAAGTACGCCACGCGTCGC-----TTTCGGCCA 211486
Qy 94  GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal 113
Db 211487 CGCGTCGATGATGCTATCCGCGAAGGCTCTGAAGCGCTGGGTATTATACCAGCCGACCAT 211546
Qy 114 SerLeu-----ThrGluLysAspGlyAlaThrThrVal-----HisIleThr 127
Db 211547 GAATTTCATCTCCGTCACCCCAAGAAAGGCGCGGAGGTATTGATCGCCAAAGTCACG 211606
Qy 128 ProGlyProArgThrLysIleAlaAsnValGlyValAlaLeuLeuGlyAspIleLeuSer 147
Db 211607 CCAGCGCTGCGGTGTTAATTTGGCGGCACCGATGTGGTATTCGCGGGGGCGCGGACC 211666
Qy 148 AspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGly 167
Db 211667 GAT-----AAAGACTATTTCGAAA-----TTGCTCGATACTCGCGCGGCTATTGGC 211711
Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 211712 ACGTACTGAACCGGCGGATTTATGAAATTTCAAAAGTCCCTTAACACAGCATTTGCGGTG 211771
Qy 188 ArgLysGlyTyr-----ProLeuAlaLysLeuGlyAsnThrArgAlaAla 202
Db 211772 CGTAAGGTTATTTCGATAGCGAATTTACCAAGCGCAGCTGGGC-----ATTGCG 211822
Qy 203 ValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIle 222
Db 211823 CTCGCGCTGCATAAAGCCTTCTGGGATATTGAT-----TATAACAGTGGCAAGCTTAC 211876
Qy 223 AlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGly 242
Db 211877 CGCTTTGGCGATGTGACCTTTGAAGGATTCACAA----ATCCCGCATGAATACGCAAA 211933
Qy 243 LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGln 262

```


sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES		Location/Qualifiers
source	1.	.10819
	/organism="Escherichia coli K12"	
	/strain="K12"	
	/sub_strain="MG1655"	
	/db_xref="taxon:83333"	
protein_bind	36.	.56
	/note="central position to predicted promoter: -73.5"	
	/bound_moiety="flvy predicted site"	
	86.	.114
	/note="factor Sigma70; predicted +1 start at 4435250"	
promoter	136.	.656
	/gene="ytfI"	
	156.	.656
	/note="b4215"	
	/gene="ytfI"	
gene	/function="orf; Unknown"	
	/note="ol66"	
	/codon_start=1	
	/transl_table=11	
	/product="orf, hypothetical protein"	
CDS	/protein_id="AAC77172.1"	
	/db_xref="GI:1790661"	
	/translation="MLPRIHNNFTGAVELFKVSSYTKTHSNFFNHHAKKKDWI SNYSLTLUREFRCATQIDKSGYQVLSSKNETVHAMDKFLISFKDKNGAETMTLR GGGFYEIEIPITINEYSPMDFKNREFPLEQTDCCMPGIFYRKNSLTYYRKLKLVIFTR RLVTCL"	
	complement(1156..1710)	
	/gene="ytfJ"	
gene	/note="b4216"	
	complement(1156..1710)	
	/gene="ytfJ"	
	/function="orf; Unknown"	
	/note="f184; 100 pct identical amino acid sequence and equal length to YTFJ_ECOLI SW: P39187"	
CDS	equal length to YTFJ_ECOLI SW: P39187"	
	/codon_start=1	
	/transl_table=11	
	/product="orf, hypothetical protein"	
	/protein_id="AAC77173.1"	
promoter	/db_xref="GI:1790662"	
	/translation="MTLRKILALTCLLLPWASAHOFQGVPPPIGIDRGELVLDK DPSYKTNASQVLGKRVLRQIAGRTSAKERNATLIEAIKSAKLPHDRYQTTIVNT DAIPGSGMFVRSLSNKKLYPWSQFIVDSNGVALGAWOLDEESSAVVVLKDRGVQ WAKDGALTPEEQVQVMDLIQKLK"	
	complement(1790..1807)	
	/note="factor Sigma54; predicted +1 start at 4436912"	
	complement(1882..1910)	
promoter	/note="factor Sigma70; predicted +1 start at 4437004"	
	1924..1951	
	/note="factor Sigma70; predicted +1 start at 4437087"	
	1996..2241	
	/gene="ytfK"	
gene	/note="b4217"	
	1996..2241	
	/gene="ytfK"	
	/function="orf; Unknown"	
	/note="o81; 100 pct identical amino acid sequence and equal length to YTFK_ECOLI SW: P39318"	
CDS	equal length to YTFK_ECOLI SW: P39318"	
	/codon_start=1	
	/transl_table=11	
	/product="orf, hypothetical protein"	
	/protein_id="AAC77174.1"	
promoter	/db_xref="GI:1790663"	
	/translation="MAPHRNTQOEITMKIFORYNPLOVAKYKILFRGRGLYIKDGA FEPDKGKILIPKVKDKLHLSVMSEVNRQVRLQTEMA"	
	complement(2320..3663)	
	/gene="ytfL"	
	/note="b4218"	

Alignment Scores:
Pred. No.: 1,11e-15
Score: 379.50
Percent Similarity: 39.46%
Best Local Similarity: 24.66%
Query Match: 11.96%

CDS	complement(2320..3663)	
	/gene="ytfL"	
	/function="putative transport: Not classified"	
	/note="f447; 100 pct identical amino acid sequence and equal length to YTF_L_ECOLI SW: P39319"	
	/codon_start=1	
promoter	/transl_table=11	
	/product="putative transport protein"	
	/protein_id="AAC77175.1"	
	/db_xref="GI:1790664"	
	/translation="MLNSILVILCLIAVSAPFMSSEISLAASRKIKKLLADGNINA IRLNQENPQGMFTVQIGLNAVAILGGVIGDAFSPAFHSLSFSRYMAEISEQLSF ILSFLVTKGMFIFLADLTQKRGIMAEPAVALRIINPMRFCLVVCPTPLVMFENGLANI IRNFKLPWVRKDDITSDDIYAVVEAGALAGLVRKOEHLIENVELESTRTVPSSMTP RENWFDLHDEDEOSLKNKVAEPHPSKFLVCNEDIDHIIIGYVDSKDLNRVLANSIA LNSGVOIRNTLIVPDTLTLSEALESKTAGEDPAVIMNEYALVVGITLINDVMTILMG DIVGGLEQIIVARDENSWLIDGGTPIDDVMRVLDIDEPQSQNYETIGTFMFMFLRK IPKRTDSVAFAGYKFEVDIDNVDQLLVTRIDSKATALSPKLPAKKEESA"	
gene	complement(3722..3750)	
	/note="factor Sigma70; predicted +1 start at 4438844"	
	complement(3785..3813)	
	/note="factor Sigma70; predicted +1 start at 4438907"	
	3905..3938	
repeat_region	/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"	
	complement(3986..4624)	
	/gene="msrA"	
	/note="b4219"	
	complement(3986..4624)	
CDS	/gene="msrA"	
	/function="enzyme; Proteins - translation and modification"	
	/note="f212; 100 pct identical amino acid sequence and equal length to PMSR_ECOLI SW: P27110"	
	/codon_start=1	
	/transl_table=11	
promoter	/product="peptide methionine sulfoxide reductase"	
	/protein_id="AAC77176.1"	
	/db_xref="GI:1790665"	
	/translation="MSLFDKKHLVSPADALPGRNTPMPVATLHAVNGHSTNVPDGM IAIFAMGCFWQVERLEWQLPQVYSTAAGYTGTPNPTNEVCSGDTGHAEAVRIVYD PSVYSYEQILOVFNENHDQGMROGNDHGTQYRSIYPLTPQDAAAARSLERFOAA MLAADDHRITTEIANATPFYAEVDDHQVYLHNKPYGYCGIGGIGVCLPPEA"	
	4662..4690	
promoter	/note="factor Sigma70; predicted +1 start at 4439826"	
	complement(4687..4718)	
	/note="factor Sigma70; predicted +1 start at 4439809"	
	4751..4762	
	/note="central position to predicted promoter:61"	
gene	/bound_moiety="Arac predicted site"	
	4830..6563	
	/gene="ytfM"	
	/note="b4220"	
	4830..6563	
CDS	/gene="ytfM"	
	/function="orf; Unknown"	
	/note="o577; 100 pct identical amino acid sequence and equal length to YTFM_ECOLI SW: P39320"	
	/codon_start=1	
	/transl_table=11	
promoter	/product="orf, hypothetical protein"	
	/protein_id="AAC77177.1"	
	/db_xref="GI:1790666"	
	/translation="MRTIROLCCVSLCLSGSAVANVRLOVBGLSGOLEKNVRAQLS TIESDEVTDRFRARVDIAIREGLKALGYQPTIEFDLRPPKPKGRQVLIATKPGV PVLIGTGVYLRGGARTDKDYLKLLDTRPAIGTVLNQGDYENFKKSLTSLTRKGYFD	
	1,11e-15	
gene	379.50	
	Matches: 145	
	Conservative: 87	
	Mismatches: 245	
	Indels: 111	

DB:	1	Gaps:	20
US-09-857-669-2 (1-609) x AE000493 (1-10819)			
Qy	74	LeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu	93
Db	4956	CTTTCTACGATTTGAAGATGATGAAGTGACGCCAGACCGTCGC-----TTTCGGCGCA	5006
Qy	94	GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal	113
Db	5007	CGCGTCGATGATGCCATCCGCGAAGGCTGAAAGCGCTGGGTATTATCCACCGCACCAT	5066
Qy	114	SerLeu-----ThrGluLysAspGlyAlaTyrThrVal-----HisIleThr	127
Db	5067	CAATTTGTATCTCCGTCACCCCAAGAAAGCGCGCAGGATTGATCGCCCAAGATCAG	5126
Qy	128	ProGlyProArgThrLysIleAlaAsnValGlyValAlaLalLeuGlyAspIleLeuSer	147
Db	5127	CCAGCGCTCGCGGTGTTAAATTTGGCGCACCGCATGTGTTATTCGCGCGCGCGGAC	5186
Qy	148	AspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGly	167
Db	5187	GAT-----AAGACTATTGAAA-----TTGCTCGATCTCGCCCGGCTATTGGC	5231
Qy	168	SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr	187
Db	5232	ACGTTACTGAACAGGCGGATATGAAATTTCAAAAGTCTTTAACCCAGCATTCGCTG	5291
Qy	188	ArgLysGlyTyr-----ProLeuAlaLysLeuGlyAsnThrArgAlaAla	202
Db	5292	CGTAAAGGTTATTTTCGATAGCGAATTTACCAAGCGCAGCTGGGC-----ATTGCG	5342
Qy	203	ValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProfile	222
Db	5343	CTCGCGCTGCATAAAGCCTCTCGGATATTCAT-----TATAACAGTGGCGAACGTTAC	5396
Qy	223	AlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGly	242
Db	5397	CGCTTTGGCATGTGACCTTTGAAGGATCACA-----ATCCGCGATGAATACCTGCAAA	5453
Qy	243	LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGln	262
Db	5454	CTGTGCGGTTTAAAGAGGCGGATGACTACGAATTCGAAGATCTGGCAGAACTGAACCG	5513
Qy	263	AlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu	282
Db	5514	CGACTTCTGCTACCGCTGTTTAACCTCGGTGGTGGTGGCTCCACAAATTTGATAACCG	5573
Qy	283	GlnGlyAspArgVal---ProValLysValSerValThrGluValLysArgHisLysLeu	301
Db	5574	CGCGAAACGAAAGTATTACCAATTCACGGCGTGGACCGCGTGAAGCGACCTGG-----	5633
Qy	302	GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr	321
Db	5634	GAACACCGGGTGGTGTACTCTACGCGATGGGACCGCGCTGGAAGCGACCTGG-----	5687
Qy	322	TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu	341
Db	5688	-----AAAAGCCGTGGATGACTCTTAT-----GGTCACAGT	5720
Qy	342	ThrThrLeuAlaLagIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer	361
Db	5721	CTGACCAACCACTACTAGTATTTCGCGCGCGCAACAG-----ACCTC	5762
Qy	362	AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly	381
Db	5763	GACTTCACGCTATAAATCCGCTCTCTGAAGAAT-----	5795
Qy	382	IleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAla	401
Db	5796	-----CCACTGGACCAATATTATTGGT	5819
Qy	402	GluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu	421
Db	5820	CAGGCG---GGTTTTAAGCGCACTGACCTGAACGATACCGAATCTGACTCCACTAGCGTG	5876
Qy	422	ThrAlaSer-----TrpLysArgGlnLeuLeuAsnVal	433
Db	5877	GTGGCTTCTTCGCTACTCGGATCTCTCCAGCGCTGGCAGCGTGCATTT-----AAC	5927
Qy	434	LeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrLeuGlyThrPhe	453
Db	5928	CTGGCTGGAGTCTCGACCACTTACTCAGGCTGAATTAACCAATACCAACGATGCTGTTT	5987
Qy	454	LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhe-----	469
Db	5988	TATCTGGGTGATGATTAGCCGCGACGCTTCTCGTGTGGCTGATGCCAACCTGGGGC	6047
Qy	469	-----	469
Db	6048	GACTCGCAACGCTACTCTCTGACTACTCTCCAACACGCGCTGGGTTCAGATGTCGATTTC	6107
Qy	470	-----PheThrProGluAsnLysLysLeuGlyThr-----PheIle	481
Db	6108	TCCGTTTTCAGGCGCGACAGACGCTCTGGATCCGCACACTGTACGATCGCCATCTGTTT	6167
Qy	482	IleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeu	501
Db	6168	ACACGCGCACGCTGGCTGGATTGAAACCGGTGATTTTCGACAAAGTACCCCGGATCTG	6227
Qy	502	MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu	521
Db	6228	CGTTTCTTCCGCGGGCGACCGCAGTATCTGCTGCTACAAATCTATCGCTCCG	6287
Qy	522	AlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGluTyr	541
Db	6288	AAATACGCCAAGCTGACCTGAAAGGCGCTCGAAGTTGTAAACCGGATCTCGGAATAC	6347
Qy	542	GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla	561
Db	6348	CAGTACAAACGTGACCGGAAATGTGGGCGCGGTGTTTGTGATAGTGGCGAAGCGGTA	6407
Qy	562	AlaAsnPheLysArgMetLysLeuHisGlySerGlyLeuGlyValArgTrpPheSer	581
Db	6408	AGCGATATTCGCCCGCAGCGACTTTAAACCGGTACCGGGTCCGCGTCCGCTGGAATTC	6467
Qy	582	ProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys-----	598
Db	6468	CCGTCGGCGCAACTCAAACTCGATTTTCCCTACCGCTCGCGGATAAAGACACACGGG	6527
Qy	599	IleArgTrpHisIleSerLeuGly	606
Db	6528	TTACAGTTTATACATCGGTCTGGGG	6551
RESULT 12			
AE005654			
LOCUS	AE005654	11760 bp	linear
DEFINITION	Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 273 of 290.		
ACCESSION	AE005654		
VERSION	AE005654.1	GI:12519212	
KEYWORDS			
SOURCE	Escherichia coli O157:H7 EDL933.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 11760)		
AUTHORS	Perna, N.F., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Rackett, J., Klink, A., Boutin, A., Shao, Y., Miller, L., Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamouglis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.		
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7		
JOURNAL	Nature 409 (6819), 529-533 (2001)		
MEDLINE	21074935		

PUBMED 11206551
REFERENCE 2 (bases 1 to 11760)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source Location/Qualifiers
1. .11760
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
126. .866
/gene="cysQ"
/note="25825"
126. .866
/gene="cysQ"
/function="phenotype: Central intermediary metabolism: Sulfur Metabolism"
/note="Residues 1 to 246 of 246 are 99.18 pct identical to residues 1 to 246 of 246 from Escherichia coli K-12 Strain MG1655: B4214"
/codon_start=1
/transl_table=11
/product="affects pool of 3'-phosphoadenosine-5'-phosphosulfate in pathway of sulfate synthesis"
/protein_id="AAG59412.1"
/db_xref="GI:12519213"
/translation="MDUQCQLARNAGDAIMQVYDGAKPMVVSKADNSPVTAAADIAA HTVIMDLRLTPDIPVLEEDPPGVEHWRWYLVDPDLOGTKFIRNGEFTYIAA ALTDHCKPTLGVVYAPVMVMYSAACKWKECGYRKQIQVDRARPPLVYSRSHAD AELKYLQOLGHEHTTSIGSLKFLCLVAEGQAGLYPFGPTNIWDTAAGHAAVAAAAGAHVHDMGKPLDYPRESFELNPGFRVSIY"
1078. .2016
/gene="25826"
1078. .2016
/gene="25826"
/function="orf; Unknown function"
/note="Residues 1 to 133 of 133 are 98.49 pct identical to residues 1 to 133 of 166 from Escherichia coli K-12 Strain MG1655: B4215"
/codon_start=1
/transl_table=11
/protein_id="AAG59413.1"
/db_xref="GI:12519214"
/translation="MLPRIRHNNFICAGVELFYKSSVTYKTHSNFFNNIIHAFKKDWI SNVDSLTREFRCATQIDKSYQVLSKNETVNAWDKPLISFLKDNAGAEYTMTLR GSGFEYEIPITINEYSFMDFLEQNRRLYAWDILQKKQSDIPKRIKGYIHH AIGDVSGLALLDDIVSKLRGFLQPGGKQCDGWYIYEKIIDDNFIVIESLIG FALKIYGGDERFNGSSVLEDEDYSLIYNFLNAGCCQVELAEQVDAIVSANLAADS DITKEICEKYKTEIAEFKQALPLVLRKKNSEI"
complement(2079. .2633)
/gene="ytfJ"
/note="25827"
complement(2079. .2633)
/gene="ytfJ"
/function="orf; Unknown function"
/note="Residues 1 to 184 of 184 are 98.91 pct identical to residues 1 to 184 of 184 from Escherichia coli K-12 Strain MG1655: B4216"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59414.1"
/db_xref="GI:12519215"
/translation="MTLRKILALTCLLLPMMASAHOFETGQRVPPIGITDRGELVLDK

gene
/gene="ytfK"
/note="25828"
2919. .3164
/gene="ytfK"
/function="orf; Unknown function"
/note="Residues 1 to 81 of 81 are 100.00 pct identical to residues 1 to 81 of 81 from Escherichia coli K-12 Strain MG1655: B4217"
/codon_start=1
/transl_table=11
/protein_id="AAG59415.1"
/db_xref="GI:12519216"
/translation="MAPHRNTQOEITMKIFQRYNPLOVAKYVKILFRGLYIKDVGA FEFDKGILIPKVKDKLHLSVSEVNRQVRMLQTEMA"
complement(3260. .4603)
/gene="ytfL"
/note="25829"
complement(3260. .4603)
/gene="ytfL"
/function="putative transport; Not classified"
/note="Residues 1 to 447 of 447 are 100.00 pct identical to residues 1 to 447 of 447 from Escherichia coli K-12 Strain MG1655: B4218"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAG59416.1"
/db_xref="GI:12519217"
/translation="MNSILVILCIIVSAFFSMSEISLAASRKIKLKLADEGINA QRVLNQENPGHEFTVQIGLNAVAIIGGIVGAARSPAFSLFSRYMSAEQLSLF ILFSLVIGMFIIFADLPTRIGIAPEAVGALRIINMRPCLVFCPLVWFFMGLANI IFRIFLWVRKDDITSDDIYAVEAGALAGVLRKHELIENVFLESRTVPSSMTP RENVIWFLDEHDEQSLKNRVAEPHFSLVCNEDIDHIIGYVDSKDLLNRVLANSOLA DLSVGQIRNTLIVPDTLTLSALESEKTEGPAVTMNEVALVVGITLNDVMTLMG LNVGGLSEQIIVARDENSLIDGGTPIDDVMRVLDIDEPQSGNYETIGCFMFMELRK IPKRTSVKFPAGYKFEVDIDNYRIDQLLVTRIDSKATSLSPKLPDAKKEESA"
complement(4926. .5564)
/gene="msrA"
/note="25830"
complement(4926. .5564)
/gene="msrA"
/function="enzyme; Macromolecule synthesis, modification: Proteins - translation and modification"
/note="Residues 1 to 212 of 212 are 99.52 pct identical to residues 1 to 212 of 212 from Escherichia coli K-12 Strain MG1655: B4219"
/codon_start=1
/transl_table=11
/product="peptide methionine sulfoxide reductase"
/protein_id="AAG59417.1"
/db_xref="GI:12519218"
/translation="MSLFQKHLVSPADALPGRNTPMVATLHVNCHSMNTVPDGM EIAIFAMCGFWGVERLWOLHGVSTAAGYTGVTNPNTYEVCSGDTGHAEEVRIYVD PSVISTEQQLQVWENHDPAQNRQGNHGTQYRSALYPLTTPQDAAASLERFOAA MLAADDDRHITTEIANPTFYAEDDQQYLHRNPYGCYGGIGGICLPPPEA"
5770. .7503
/gene="ytfM"
/note="25831"
5770. .7503
/gene="ytfM"
/function="orf; Unknown function"
/note="Residues 1 to 577 of 577 are 100.00 pct identical to residues 1 to 577 of 577 from Escherichia coli K-12 Strain MG1655: B4220"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59418.1"
/db_xref="GI:12519219"


```

Db 7063 CAGAACGTCGTGGATCCGCACACTGTACGATCGCATCGTTTGTGACACGCGGCACGCTG 7122
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 487 GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7123 GGCTGGATTGAACCGGTGATTTCGACAAAGTACCGCGGATCGCTTTCGCGCGG 7182
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 507 GlyAlaSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGly 526
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7183 GGCATCGCATATTCGCGGTACAAATACAAATCTATCCGCCGGAATACGTTACGGT 7242
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 527 SerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThr 546
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7243 GACCTGAAGGGCGCTCGAAGTTGATAACCGGATCGCTGGAATACCACTACACGTGACC 7302
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 547 ArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArg 566
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7303 GGAATATGGTGGCGCGGTGTTGTGATAGTGGCGAAGCGGTAAAGATATTCGCCGC 7362
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 567 MetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe 586
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7363 AGCGACTTTAAACCGGTACCGGGTGGCGGTACGCTGGGAATCGCGGTGGCGCAATC 7422
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 587 SerPheAspIleAlaTyrGlyHisSerAspLysLys-----IleArgTrpHisIle 603
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7423 AAATCGATTTCGCGGTACCGGTGGCGGATAAAGACACACACGGGTTACAGTTTATCATC 7482
      :::::      :::::      :::::      :::::      :::::      :::::
Qy 604 SerLeuGly 606
      :::::      :::::      :::::      :::::      :::::      :::::
Db 7483 GGTCTGGGG 7491

RESULT 13
ECOWU93      338534 bp      DNA      linear      BCT 30-JAN-2001
LOCUS      Escherichia coli K-12 chromosomal region from 92.8 to 90.1 minutes.
DEFINITION      U14003
ACCESSION      U14003.1 GI:1263172
VERSION      1
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 338534)
Burland,V., Plunkett,G. III, Sofia,H.J., Daniels,D.L. and
Blattner,F.R.
Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes
Nucleic Acids Res. 23 (12), 2105-2119 (1995)
95334362
7610040
2 (bases 1 to 338534)
Plunkett,G.I.I.I.
Direct Submission
TITLE      Submitted (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics,
JOURNAL      University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
MEDLINE      Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
PUBMED      608-263-7459
REFERENCE      On Apr 17, 1996 this sequence version replaced gi:536929.
AUTHORS      This sequence was determined as part of the E. coli Genome Project
      (Frederick R. Blattner, director) at the University of
      Wisconsin-Madison. Supported by award HG00301 from the NIH Human
      Genome Project. The entire sequence was independently determined
      from E. coli MG1655; overlaps with other sequence determinations
      are annotated. The start of this entry overlaps the end of the
      entry ECOWU99 (000006) by 1885 bp.
      Location/Qualifiers
FEATURES      1. 338534
source      /organism="Escherichia coli"
      /strain="K-12"
      /sub_strain="MG1655"
      /db_xref="taxon:562"
      /map="92.8 to 100 minutes"
      /note="This sequence comprises the following lambda

```

```

clones: DD928(EC27-278), DD930(EC30-188),
DD933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070),
DD941(EC17-142), DD945(EC17-8), DD947(EC24A-34),
DD949(EC22-169), DD952(EC27-297), DD953(EC27-409),
DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2),
DD962(EC21-104), DD965(EC23A-40), DD968(EC30K660A-4pp),
DD970(EC19-202), DD974(EC19-61), DD975(EC18-233),
DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3),
DD984(EC17-101), DD987(EC18-115), DD990(EC17-136),
DD992(EC18-282), DD995(EC17-115), DD997(EC18-113); M13mp19
or Janus vectors were used for subcloning"
1. 1885
/notes="1885 bp overlap with end of GenBank Accession
Number U00006 (ECOWU89)"
complement (<1. 398)
/notes="ORF_f326 of GenBank Accession Number U00006
(ECOWU89)"
/codon_start=1
/transl_table=11
/label_ORF_f132p
/protein_id="AAA96985.1"
/db_xref="GI:1263173"
/translation="MGFTTRVKSESEKPFNFALFDWKYGTFFILAIIVAFGSLSP
EYFLTNITQIFVQSVTVLIGMGFEFFAILVAGIDLSVGAIALSGMVTAKLMLAGV
DPLAAIMIGVLVGGALGAINGLVNWTL"
277. 397
/notes="predicted bend of 75.10 degrees"
complement (377..1909)
/notes="ORF_f510"
/codon_start=1
/transl_table=11
/label_ORF_f510
/protein_id="AAA96986.1"
/db_xref="GI:536931"
/translation="MATPIISMAGIKSGFQPVHALKSVNTVTPGEIHALGNGAGK
STMLKVLGCIHEPTKGTITINNYSYNKLDHKLAAQDGIIGYBELSIDELTVLENY
IGRHLTKGICGVNIIDWRENRVAAAMLLRVGLVDLDEKVNLSHSHQMLEIAKTL
MLDQVITMDEPTSSLTKEVDYLFILMOLKEGTIAIVYISHKLAIRICDRTVTM
KDGSSVCSGIVSDVNDIDVRLMVGRELONRFNMKNVSNLAHETVFEVNRVTSRDR
KVRDIFSVCERGEILGFAGLVSGRTLMNCLFGVDKRRAGGEIRLNGKDISPSPLD
AVKKGMAYITESRRDNGFFPNFSIAQNMALSRSLKDGKYGKAGMLFHEVDEQRTAEHQ
RELLALKCHSVNQNTITELSGGNQKQVLIISKWLCCPEVIIFDEPTGIDVGAKEIYK
VMRLADDDGKVLWVMSSELPIITVCDRIAVFCGEGLTQILTNRDDMSSEEMAWALP
QE"
809. 931
/notes="predicted bend of 74.75 degrees"
1788. 1889
/notes="predicted bend of 75.87 degrees"
complement (1983..2010)
/notes="promoter matrix score of 55; putative; resembles
flh-regulated promoter"
complement (1985..2021)
/standard_name="REP; repetitive extragenic palindromic
element"
/notes="contains 1 REP sequence"
complement (2036..2971)
/notes="ORF_f311"
/codon_start=1
/transl_table=11
/label_ORF_f311
/protein_id="AAA96987.1"
/db_xref="GI:536932"
/translation="MNKYLKVFSGTLVGLMLSTAFAPAAAEVAVVVKLTLSNPNVDMKK
GIEDEAKTLGVSVDIFAPSGEGFQSQQLQFEDLSNKNYKGIAPLSSVNLVMPYAR
AWKKGIYLNDEKIDMDNLKAGGNVEAFVTDNVAVGARGASFIIDKLGAEGEVNA
IIEGKAGNASGEARRNGATEAFKXAKIKLVAQPADMDRIKALDVATYQIRNPNIK
AIYCDMTMAEGVAQAVANAGKTGKLVVGTDGIPKARKVMEAGQMTATVAQNPADIG
ATGLKLMVDAEKGSKGVIPLDKAPEFKLVDSILVTQ"
3010. 3121
/notes="predicted bend of 73.22 degrees"
complement (3030..3953)
/notes="ORF_f307"
/codon_start=1

```


RESULT 14

```
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAsp 217
Db 499 ATTTTCAGCTTTAGAAATCCCGCTGAACCCATCAAGCATGGTGGCGAAATGTTATTCAT 558
Qy 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
Db 559 AGTGGTGTCCGTTATCATATATGGAATATATACCTTTTACCCATTCACAA---ATTCCGGAT 615
Qy 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
Db 616 GATTACTTAATAATATCTTAACATCAATCTGGTATCGGTATTTAATGAATAATTTG 675
Qy 258 LeuAspPheGlnGlnAlaLeuGlnGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
Db 676 TCGGATTTAACACGAGATTTTCCATCTCAAAATTTGGTTAGCTCAGTATTAGTTACGCT 735
Qy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
Db 736 AATGTTAATCAT---AAAGCAAACTGTGGATGTGGAGATTATCTTTATCCACGTAAA 792
Qy 298 ArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIle 317
Db 793 AAAAATCGATGGAACCTCGGTGGCTTTTCTACTCATGCGCGCTTCACGGACAATA 852
Qy 318 AlaTyr----- 319
Db 853 GCCTGGCAAAACCTTGGATTAATAGCGGTGGACATAGTTTGGCTTCAAAATCTTTATCTC 912
Qy 319 ----- 319
Db 913 TCTCCACCAAAACAACTCTAGAGGCAACTTATCGAATGCCACTGCTTAAAAATCCATTA 972
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLys 339
Db 973 AATTATTACTATGATTTTGGCGTGGTGGGAAGGGGAAAGAGATGATACCAATACG 1032
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
Db 1033 AGACTGCTTAGCTTGTGACGC-----TTACGTTATTTGG 1065
Qy 360 ThrSer-----AsnValSerTyrAsnArgSerThr 369
Db 1066 AATAATCGCATGCTGGCAATATTTTGGCGGACTTCTGTATGCCATACGACGTTTATACA 1125
Qy 370 ThrGlnAsnLeuGluLysArgAlaPhe-----SerGlyGlyIleTrpTyrVal 385
Db 1126 CAAGCGGATATCACTGATAAAACCTTACTTCTTTATCCAACTGTTGCATTTACTCGTACT 1185
Qy 386 ArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLys 405
Db 1186 CGATTACGTGGTGGTTCCTTTGCCACTTGGGCG-----GATGTGCAAAA 1230
Qy 406 IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrp 425
Db 1231 ATT-----ACTTTTGATTTAAGC----- 1248
Qy 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1249 AAACGA----- 1254
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1255 -----ATTGGCTATCAGAATCTCTTTTATATAAAGTGCAGACATCT 1296
Qy 466 AlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
Db 1297 AGCCGCTGGGTCTTACTTATGCAAGAAATCATCTGCTC-----GTTGCTCGTGCT 1347
Qy 485 GlnAlaGlyTyrThrValAlaAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 1348 GAAATCGGGTATTATACATACAAAAGGTATTGAAAAAATTCGCGCTACATCGGTTCTTT 1407
Qy 505 SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
```

```
Db 1408 GCTGTGGCATCGTAGTGTGCGCGTTACGGCTATATAAAAAATTCGCCCTAAAAATAGA 1467
Qy 525 AsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro 544
Db 1468 AATGGAATAATGGTGGTGGTTCGCGATTGCTTACCACCTCTTTTAGAATATATCAATACAA 1527
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhe 564
Db 1528 GTCTATCCAAATGGTGGCGGCAACTTTTCACATAGTAGTGGATAGCTGGCGATAATTAC 1587
Qy 565 LysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 1588 ACAGCAAAAGAGCTGCGTTATGTCACAGGCGTGGTGGCTGGCATGCCAGTTGGT 1647
Qy 585 ProPheSerPheAspIleAlaTyrGlyHisSerAsp-----LysLysIleArg 600
Db 1648 GCGATTAAATTTGATTTGGCCACACCCTTGGTGATTAAGATAACAGCAAAAATATTCAA 1707
Qy 601 TrpHisIleSerLeuGlyThr 607
Db 1708 TTTTACATCGGACTTGTGACA 1728
```

RESULT 15
U32752/c

LOCUS U32752 11833 bp DNA linear BCT 29-MAY-1998
DEFINITION Haemophilus influenzae Rd section 67 of 163 of the complete genome.
ACCESSION U32752 L42023
VERSION U32752.1 GI:1573692

KEYWORDS

SOURCE

ORGANISM

Haemophilus influenzae Rd.
Haemophilus influenzae Rd.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus
1 (bases 1 to 11833)
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,
Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,
Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,
Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrman, J.L.,
Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M.,
Smith, H.O. and Venter, J.C.
Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

REFERENCE

5 (bases 1 to 11833)
 White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D., Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 REMARK The whole genome was shifted by 588 nucleotides for a new start
 COMMENT On Sep 30, 1996 this sequence version replaced gi:1221393.
 FEATURES
 Location/Qualifiers
 1..11833

source

/organism="Haemophilus influenzae Rd"
 /db_xref="taxon:71421"

gene

145..939

CDS

/gene="HI0690"

145..939

/gene="HI0690"

/note="similar to GB:LL19201 SP:P11244 GB:M55990 GB:U13915
 GB:X15034 percent identity: 73.18; identified by sequence
 similarity; putative"

/codon_start=1

/transl_table=1

/product="glycerol uptake facilitator protein (glpF)"

/protein_id="AAC22350.1"

/db_xref="GI:1573694"

/translation="MDKSLKNCIGEFGLTALLIFFGVCVAAKLVAGASFGLEWFEISI
 MNGMGLAVATAGLSGHLNPAVTIALMKFACDGRKVPYIIISQMLGAPFAAALV
 VALYRNVDIVETVNIIVRGTOESLSLGTSTYPPHSLISGGAFAVEFTAILMAL
 IMALEDGNGVPRGPLALLIGILLIAYIGGANGPLTGAMNPARDFGPKFFAYLAGWG
 ELALTGGREPIPIFVPMVAPVLGALAGAWLYKKAIGGNLPCNCGCE"

960..2471

/gene="HI0691"

960..2471

/gene="HI0691"

/note="similar to GB:LL19201 SP:P08859 GB:M18393 GB:M55990
 GB:X15034 percent identity: 76.89; identified by sequence
 similarity; putative"

/codon_start=1

/transl_table=1

/product="glycerol kinase (glpK)"

/protein_id="AAC22351.1"

/db_xref="GI:1573695"

/translation="MTDKKYLIALDQGTSSRAVLDDHNNVVEIAQREFTQIYPRAG
 WYHEPMETWATQSTLNEVAKAGITDEIAAIGITNORETIVWEKSTGTPVYNAI
 VWCQRADITDKLADGHEEYIRNTGLVDPVPSGTRKVKWILONDEKAREGE
 LLEFGVDVTLWKLQGRVHVTDYTNASTMLENHTKOWDDKMLEILNIPRSMLEPV
 RNSSEIYQWLTGGGGVRLPVAGIAGDQQAALYHLCVHQAQNTYCTGCPMLLHT
 GKNAITKGLLTITACNAKGEPEYALGESVFIAGASQWLRLKIVHDSFDSYFA
 QKVTDNSGVYVPAFTGLGAPYMDPYARGAIFGLSRGANNRHIVRATLESIAQTRDA
 LEAMQSDGRLQYLRVGGATNNFLMQFQADILDVNVPRVVEVTALGAAYLAGL
 ATCPWKDLDELKARVERTFSPDSNKRERRRYGKWKAVKRSLEWAKEDDEE"

complement(2546..3013)

/gene="HI0692"

complement(2546..3013)

/gene="HI0692"

/note="similar to SP:P26972 GB:X52093 GB:X63336 PID:47695
 percent identity: 74.03; identified by sequence
 similarity; putative"

/codon_start=1

/transl_table=1

/product="xanthine-guanine phosphoribosyltransferase
 (gptB)"

/protein_id="AAC22352.1"

/db_xref="GI:1573693"

/translation="MSERYVTWDMFQMHARLSERLLPASQWKGLIIVSRGGLFPAA
 VLARELGHFEIVTCIASYHDHNNQGEQLVHLAAQVPGNGEGFIVVDDLDVDTGNARA
 IQMTPNARKEVTVFAKPAAGAEVDVVDIPQNTWTEQPMDLGLTFVPLSRK"

3297..4121

/gene="HI0693"

3297..4121

/gene="HI0693"

/note="similar to GB:M68502 SP:P26093 PID:1573696 percent
 identity: 99.64; identified by sequence similarity;
 putative"

/codon_start=1

/transl_table=1

/product="lipoprotein E (hel)"

/protein_id="AAC22353.1"

/db_xref="GI:1573696"

/translation="MKTTLKMTALAALSAFVLAGCGSHMKSEGHANMLQQAVALGL
 NMWQSGEYKALAYQAYNAAKVAFDAKVKKAVVADLDLDETHLSDSPYAGWQVN
 NKPFQDKWTRVDARQSRVPGAVGFNNYVNSHGKVFYVYVTRKDSYKSTKIDDMK
 RLGPNGVESAFYLLKKDKSAKAAFEATEKQCYEIVLVXGDNLDGFGNTVYGLKNADR
 RAFVDONOGKFGKTFIMLPNANYGMEGSLABGYFKKDTQGOIKARLDAVQAWDK"

4261..4983

/gene="HI0694"

4261..4983

/gene="HI0694"

/note="similar to GB:U00096 SP:P75966 PID:1787380 percent
 identity: 66.29; identified by sequence similarity;
 putative"

/codon_start=1

/transl_table=1

/product="conserved hypothetical protein"

/protein_id="AAC22354.1"

/db_xref="GI:1573697"

/translation="MMJQAFQNRNHLKMKNFSAKLPKSGKSAVNHFHTFKPKLKPK
 TSLDLETWLVFNKPDFVLTQFTDEQGRATLKDFTSIPNVYAAGRLDSEGLLITN
 NGELQRLADPKFTEKTYWQVEGIPETDLAQLRKGVKELDGVTKSAKVRLLISEPN
 LWERNPPIRERKNITPSWLEIKISEGRNQVRMTAHIGFPTLRLVRVSMGLLSINGL
 ENGSPRLLSLDEIKALFOTVKI"

complement(5109..6080)

/gene="HI0695"

complement(5109..6080)

/gene="HI0695"

/note="similar to GB:L06129 SP:P29014 PID:147343 GB:U00096
 PID:1788848 percent identity: 53.62; identified by
 sequence similarity; putative"

/codon_start=1

/transl_table=1

/product="exopolyphosphatase, putative"

/protein_id="AAC22355.1"

/db_xref="GI:1573698"

/translation="MNSDILEPKHRGNVREIAAIDLGSNSFHMIVARIVNGSIOVLRS
 LKQVKVLAEGDENAVLNOEAITRGVNCIALFAERLQGEPMENNVNYYTTRRVN
 DFLRQAKRFPYPIINISGOEAKTIYAGVCHTQPEKGRKLVIDVGGSTEMIIGDD
 FTPLMAESHGMCVSFATQFTDGIISPENFORQSAVNKIEDLGLEYRKLGRQSVL
 GSSGTIKTVAOVIAITNDPNTGTITAEARNALIEOTLOAKHFTELNINGLNQDRVDV
 PGLAILSAVEDVPHIQOMRYSDGALREGVIYSLEKNFQVADIRASTA"

complement(6088..9984)

/gene="HI0696"

complement(6088..9984)

/gene="HI0696"

/note="similar to GB:U14003 SP:P39322 PID:537063 GB:U00096
 percent identity: 34.87; identified by sequence
 similarity; putative"

/codon_start=1

/transl_table=1

/product="conserved hypothetical protein"

/protein_id="AAC22356.1"

/db_xref="GI:1573699"

/translation="MTEQIQPSETSPKSPKNNKHHVRKAVCIGSAVILVPVLGVAG
 ALSFDAGQKSLIOLVDKMLDSFSEVEGQLQGLVLRVRYOTAGIETHIAQARLOL
 DFGCLIFSEVCLRDFTLNKPTIAINTALLPPSPADNSKSGMKRISLPTISINENLVM
 ODLNVINDIOTITLGNFKSAVSLNNEKGTIATPTEINDISVIAKLSKSEKSEPKABQ
 NKPVDMAATEQSIITPAFLGNVSEIILPFDLHTEISGKNWOYQAVNEKETLQSVBMS
 SLIAQADVDYDNLQLOQLAVESLNLSSQGLQDLDPLDLTLKSHLEPLKSDGKE
 IIPASDVIDITLSSGLKSTALSLKGVLDALNGVNLQADKMPUNLNLNVAKGQIT
 FYNTPLEKINDVTLLKTDGLLNYHAELKGVAGMNYIPASQVPMATLGLKHSRFGAG
 LGTSDLDGSEFVGNANNGANWDIADLERMNIATFFVPMATLGLKHSRFGAG
 OGWOVEVPVADLNGMLSAKPSLKGSAITNLQVLLTPDLQIKYSGNYLKGSLRGLDDH
 SDPALDINAPNLRLGLWSDLKGRVGRVATISGQITTPNLDLTSSNLHQLGGLAKAS
 IKGHINNASLSSGKLNIAKQOLHYGNKIKHLDDLDSGDEQNHKILLKSGQEPVAA
 LQINGHFDRLQWKGTTISQVAFETPIDGVKSNOAIVASDNKQTOQNTASHCWQNTD
 VELCFQEPNAGKQGNIPFOFVRVNLDLVNLKFEQNSLKGNLQVQNAWFFDKPQF
 TANVDGNHLAFQKLDYRTFKLYIPKLTNLADIQNNNLVLTIDINYNQGRIVGDIHL
 NDLAKNRQLGGTILAIERLNLSTANQLTSGESVNGEWSKLSFGGNLEKPLLLGDPNI

RNIETKLKSPVNIITDGDIALRPNDRNSTLOGKIKTVDSHLNLTGRANWANIIEHWTE
 LNAQNFNDIFSMKLRSPNITIKANPKELNSCTVDIPWARIKIDSLSDPTAEPV
 SEDEVNFNPHKSKKEELIKREFAAKTSGMEIRSDLRINIGKQVSDAYGLKTNLDGL
 LSVKQDKNGLGQINLTGRVASFQDILLIRKGLISFQCATQPTLNEIARNPET
 MEDSKITAGVRVIGIADSPVITFSEPKQDQALSYLLTGRSLESSEGVSGSVGA
 ALIGLITSKSKLGVSGIGEVGIGIDLMGTSVGCDKSKVTSGNITNRLQIKYGVGLF

Alignment Scores:

Pred. No.: 2,93e-15 Length: 11833
 Score: 374.00 Matches: 153
 Percent Similarity: 36.88% Conservative: 93
 Best Local Similarity: 22.94% Mismatches: 263
 Query Match: 11.79% Indels: 158
 Gaps: 18

US-09-857-669-2 (1-609) x U32752 (1-11833)

Qy 3 IleLysProThrAlaLeuLeuProAlaLeuPhePhePheProHisAlaTyrAlaPro 22
 Db 11715 CTAAACTCAGCGGCTTTTAA---GCATTAGTTCCTTCCCGCATTTGCTGAACAA 11659
 Qy 23 AlaAlaLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerPro 42
 Db 11658 ACCGTTCATATTGAAGTTCAGGCGATCGTGTTCCTGCTGCTGCGTAATACAGAT--- 11602
 Qy 43 AspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSer 62
 Db 11601 -----CTCAATGTTAATTTCGATTAAATAA 11578
 Qy 63 GluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluVal 82
 Db 11577 GAAGAAATGATGCTCGAAGCGTATCAACATTGGTTAAACCAAGCC----- 11530
 Qy 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
 Db 11529 -----GTGGATCGTGGT 11518
 Qy 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAsp----- 119
 Db 11517 TTGGCTGTGTTGTTATTATGAATCTTCCGTCGCTTTTGAACCAAAACACGCGTCAAGGC 11458
 Qy 120 -----GlyAlaTyrThrValHisIleThrProGlyThrProArgThrLysIleAlaAsnVal 137
 Db 11457 AAACGCGATTTATTGATGCTCATGTTACACCGAGCGAGCCAAACAAATTTGCGGGGACT 11398
 Qy 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
 Db 11397 GATGTCAAATTTGAGGGGAAGCGCACAGATCAAAATTT-----AAT 11353
 Qy 158 AlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
 Db 11352 GCATTGGCTAAACAACTTACCGAAAGATGGCGTTTGGTTGAGCACCAAACTTAGCATGAT 11293
 Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
 Db 11292 TACAAACAGCGGATTTCCAGTTTACGATTAATCGTGGTATTTTGTGAGGGAACCTTTAA 11233
 Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAsp 217
 Db 11232 ATTTTCAGCTTTAGAAATCAGCCCTGAAACCCATCAACGATGCTGCGCAATGTTATTGAT 11173
 Qy 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 Db 11172 AGTGGTGTCCGCTTATCATATTATGCAATATTACTTTAGCCATTACAA---ATTCCGGAT 11116
 Qy 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
 Db 11115 GATTACTTTAAATATCTTTAAACATCAAAATCTGCGTATCGGTTAATTAATGAATAATTG 11056
 Qy 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
 Db 11055 TCGGATTTAACCCAGCGATTTTCCATCTTCAAAATTTGGTTAGCTCAGTATTAGTTACGCT 10996

Qy 278 AspPheAspArgLeuGlnGlnGlyAspArgValProValLysValSerValThrGluValLys 297
 Db 10995 AATGTTAATCAT---AAAACAAACCTGCTGGATGGAATTAATCTTATCCAGTAAA 10939
 Qy 298 ArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIle 317
 Db 10938 AAAAATCGGATGGAACCTGGGTGGCTTTTCTACTGATGCGGCGTTCACGGACAATA 10879
 Qy 318 AlaTyr----- 319
 Db 10878 GGCTGGACAAACCTTGGATTAAATAGCCGTGGACATAGTTTGGCTTCAATCTTTATCTC 10819
 Qy 319 ----- 319
 Db 10818 TCTGCACCAAAACAACTCTAGAGGCAACTTATCGAATGCCACTGCTTAAAAATCCATTA 10759
 Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
 Db 10758 AATTATTACTATGATTTTGGCGTGGTGGGAAGGGGAAAAGAGAATGATACCAATACG 10699
 Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
 Db 10698 AGACTGCTTACGTTGTACGCG-----TTACGTTATTGG 10666
 Qy 360 ThrSer-----AsnValSerTyrAsnArgSerThr 369
 Db 10665 AATAATGCGCATGCTTGGCAATATTTTGGCGGACTTGTGATGCGATACGACAGTTTTACA 10606
 Qy 370 ThrGlnAsnLeuGluLysArgAlaPhe-----SerGlyGlyIleTrpTyrVal 385
 Db 10605 CAACGCGATATCACTGATATAAACCTTACTTCTTTATCCAACTGTTGGATTTACTCGTACT 10546
 Qy 386 ArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluArgLys 405
 Db 10545 CGATTACGTGGTGTCTTCTTCCACTTGGCGC-----GATGTCACAAA 10501
 Qy 406 IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrp 425
 Db 10500 ATT-----ACTTTTGATTAAAG----- 10483
 Qy 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
 Db 10482 AAACGA----- 10477
 Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
 Db 10476 -----ATTGGCTATCAGAACTCTCTTTTATAAAGTCCAGCATCT 10435
 Qy 466 AlaGlyTyrPhePheThr-----ProGluAsnLysLysLeuGlyThrPheIleArgGly 484
 Db 10434 AGCGGTGGTTCGTACTTATGTCAGAAAATCATGCTGTC-----GTTGCTCGTGT 10384
 Qy 485 GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
 Db 10383 GAAATCGGATTTTACATACAAAGGTATTGAAAAAATTCGCGCTACACTGCGTTCCTTT 10324
 Qy 505 SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
 Db 10323 GCTGCTGGCGATCTAGTGTGCGGGTTACGGCTATATAAATAATTCGCGCTTAAAAATAGA 10264
 Qy 525 AsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro 544
 Db 10263 AATGAAAAATTTGGTGGTGGTTCGCGATTCCTTACCACCTTCTTTAGAAATATCAATATCAA 10204
 Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhe 564
 Db 10203 GTCTATCCAAATTTGGTGGCGGCAACTTTTTCAGATAGTGGATAGTGGCGGATTAATAC 10144
 Qy 565 LysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
 Db 10143 ACACAAAAGAGCTCGCTTATGCGCAGAGCGTGTGCTGGCTGGCGATCCCGCATTTGGT 10084
 Qy 585 ProPheSerPheAspIleAlaTyrGlyHisSerAsp-----LysLysIleArg 600


```

Db 10083 GCGATTAAATTTGATATTCACACCCCATTCGTCATAAAGATAACAGCAAAATATTCAA 10024
QY 601 TrpHisIleSerLeuGlyThr 607
Db 10023 TTTTACATCGGACTTGGTACA 10003

RESULT 16
LOCUS AE013668 10831 bp DNA linear BCT 26-JUL-2002
DEFINITION Yersinia pestis KIM section 68 of 415 of the complete genome.
ACCESSION AE013668 AE009952
VERSION AE013668.1 GI:21957368
KEYWORDS
SOURCE Yersinia pestis KIM.
ORGANISM Yersinia pestis KIM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
1 (bases 1 to 10831)
Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Genome Sequence of Yersinia pestis KIM
J. Bacteriol. 184 (16), 4601-4611 (2002)
12142430
2 (bases 1 to 10831)
Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
Blattner,F.R. and Perry,R.D.
Direct Submission
Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. .10831
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
247. .990
/genes="cysQ"
/notes="y0654"
247. .990
/genes="cysQ"
/function="phenotype; central intermediary metabolism:
Sulfur metabolism"
/notes="affects pool of
3'-phosphoadenosine-5'-phosphosulfate in pathway of
sulfite synthesis; residues 2 to 247 of 247 are 78.45 pct
identical to residues 1 to 246 of 246 from E. coli K12 :
B4214; residues 2 to 247 of 247 are 79.26 pct identical to
residues 1 to 246 of 246 from GenPept : >gb|AAL23224.1|
(AE008906) affects pool of
3'-phosphoadenosine-5'-phosphosulfate in pathway of
sulfite synthesis (Salmonella typhimurium LT2)"
/codon_start=1
/transl_table=11
/product="sulfite synthesis pathway protein"
/protein_id="AAM84242.1"
/db_xref="GI:21957369"
/translation="MMLEIOGOLAREAGVAIMAVYQDGKPLDIAHKKDSPTVAADLA
AQHIKAGLARLPDIPLLSEEDPPANDYRQHWYLVDPDLDGTFELNRNGEFTVN
IALIDGCEPVLGVVYPTVEVMTSAANGAAKECCRRMOIQVDRQAPLPLVYVSRHS
DAELEDYLSQAGEHOTISYGSLSKFLCLVAEGKAQYLPFRGPTNVNDTAAGHAVATAAG
AQVHDWQGTLSYTPRESFNPGRVSLP"
complement(993. .1631)
/genes="y0655"
complement(993. .1631)
/notes="y0655"
/notes="residues 26 to 211 of 212 are 59.13 pct identical
to residues 1 to 184 of 184 from E. coli K12 : B4216;

residues 26 to 212 of 212 are 60.42 pct identical to
residues 1 to 185 of 185 from GenPept : >emb|CAD06884.1|
(AL627283) putative exported protein [Salmonella enterica
subsp. enterica serovar typhi]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM84243.1"
/db_xref="GI:21957370"
/translation="MKNLPTFRGKCFSTLSRISLQGHNMKNKSLLIISLITPPLVYS
AINIQNQRPVPPVGRDKGELGDDTKTNQFNKNMNSSQLPKTRVIRHQIAGRTSAKE
MNAPLDAIKQANLPREYQTTIVNTDDAIIIGTAMFVNSIESSRKEPFWQSFITDS
KGNVAKAWGVTEKSSAIVVLDPOGNNKFKVDGALTSAEVQVQVIDQLHQLVQQ"
1892. .2104
/genes="y0656"
1892. .2104
/genes="y0656"
/notes="residues 1 to 66 of 70 are 77.27 pct identical to
residues 14 to 79 of 81 from E. coli K12 : B4217; residues
1 to 66 of 70 are 75.75 pct identical to residues 23 to 88
of 90 from GenPept : >gb|AAL23226.1| (AE008906) putative
cytoplasmic protein [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM84244.1"
/db_xref="GI:21957371"
/translation="MRIFORYNPAKVYVVKTLFRGLRYIKDMGAPEPDNGKILVPKV
KKRHFVEVMAEYNRQVLRQADTLLT"
complement(2212. .3543)
/genes="y0657"
complement(2212. .3543)
/genes="y0657"
/function="putative transport"
/notes="residues 1 to 437 of 443 are 82.15 pct identical to
residues 1 to 437 of 447 from E. coli K12 : B4218;
residues 1 to 443 of 443 are 81.39 pct identical to
residues 1 to 446 of 447 from GenPept : >gb|AAL23227.1|
(AE008906) putative hemolysin-related protein [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAM84245.1"
/db_xref="GI:21957372"
/translation="MLNSILLLFLIATVSAFSLSEISLAASRKKIKLLADEGDTNA
LRVCLKQETPGMFTVQVIGLNAVALIGVIGDAFSPSKLVFVRMAPELADQASF
VGSFVLVSLFLADLTPKRGIMISPEAVAVRIVNPMRFLMIFRPLVMFNGSMNL
IFRLFKLPMVRNDDITSDDIYAVEAGALAGVLRKQEHLEIENFLESRTVPSSMTS
RNMVYFDLRESEDSIKDKISTHPSKFLVCDGHIDQVGVYVDSKDLLNRVLGNOSLV
LSGVQIRSAALVDPDTLISEALESEFTAGEDPAVLNEYALVVGIIITLNDVMTTLMG
DLVGQGEQIVARDENSWLIEGGTPIEDVMRVLIHDDPPQSGNYETIGGFMMYMLRK
IPKRTDFVYAGYKFEVDIDSYDKIDQLLVLRSLSDQAPILPKAPHESDA"
complement(3821. .4459)
/genes="msrA"
/notes="y0658"
complement(3821. .4459)
/genes="msrA"
/function="enzyme; proteins - translation and
modification"
/notes="residues 1 to 210 of 212 are 73.33 pct identical to
residues 1 to 210 of 212 from E. coli K12 : B4219;
residues 1 to 211 of 212 are 72.51 pct identical to
residues 1 to 211 of 212 from GenPept : >gb|AAL23228.1|
(AE008906) peptide methionine sulfoxide reductase
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="peptide methionine sulfoxide reductase"
/protein_id="AAM84246.1"
/db_xref="GI:21957373"
/translation="MONVNTAVIDAANALPGRLTSTIPVSPPLHAVHGHSMYIPEGMD
LAFFAMGCFWGAERLFWQPGVYSTAAGYSGGHTPNPTYHEVCSGHAEVVRVVFV

```



```
QY 349 SerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 5595 CGTAGCCT-----TGGATCAACTCCTTCGGGCATAGCTTAACC 5633
QY 369 ThrThrGlnAsnLeu-----GluLysArgAlaPheSerGlyGlyLeuTrpThr 384
Db 5634 ACCACTACCGCCTTCGGCACCTGAGCAGACATGGATTTCAGCTACCGTATCTCTTTA 5693
QY 385 ValArgAspArgAlaGlyLeuAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly--- 403
Db 5694 TTGAGAAC-----CCCTCGACGAGTATACCTGATGATACGGGGC 5735
QY 404 ---ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet----- 420
Db 5736 TTTAGACGT-----ACCGACTAAACGACACTAATTCAGATACACGACCTTGAAC 5786
QY 421 -----LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
Db 5787 GTTGACGCTTTTGGGATCTCTCCAGCGGCTGGCAAGGGCGATT-----AACCTG 5837
QY 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
Db 5838 CGTTGAGCCTCGATCATTACCCAGCAGCAGTACCGATACACGATGTGCTGTAT 5897
QY 455 SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr----- 468
Db 5898 CCGGGGTGAGTATTATCGTACCCGCCAACGTTGGTGGCGCATCGGCTCTGGGGGAT 5957
QY 468 ----- 468
Db 5958 AGCCAGCGCTATTCTATTGATGCTGTATAGCACTTGGGGTTCGGATCTTATTGTCG 6017
QY 469 PhePheThrProGluAsn-----LysLysLeuGly-----ThrPheIleIle 482
Db 6018 ATCTTTACGGCCGAGCAAGTGTGGATCCGGACCTTAGGTGAGAAAATCGTTTGTGGCG 6077
QY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
Db 6078 CGTGCAATGTGGGTGGATCGCAAAACCAATAATTTTGACCGTTCACCGCTCGCTGCT 6137
QY 503 PheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 6138 TTCTTCGGCGGGGCGACCGCAGTATCCGTCGTATAAATTCGGTAAATTTCCACCGCT 6197
QY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
Db 6198 GATAGTCAGGGAATAATTAACCGGGCTTCCAGTGTGGCAACGGCTCTCTTGAATATCAA 6257
QY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaLa 562
Db 6258 TATACTTCCAGCGTAGATGGTGGGGCGCGTTTTCGTGATCGGGCGGCGGTGAAT 6317
QY 563 AsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
Db 6318 AATTTCCAGCAAGAGTGACTTAAACACAGGGCAGGTGTCGGGTAGCTTCGGCATCACCC 6377
QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys-----Ile 599
Db 6378 GTTGGACCATCAAACTGATCGCGGCACCAATAGCGGCAACGACACATCGCGGTG 6437
QY 600 ArgTrpHisIleSerLeuGly 606
Db 6438 CAATTTTACATCGTTTGGGG 6458
```

```
RESULT 17
AJ414157/c AJ414157 216050 bp DNA linear BCT 06-JUN-2002
LOCUS Versinia pestis strain CO92 complete genome; segment 17/20.
DEFINITION AJ414157 AL590842
ACCESSION AJ414157.1 GI:15981328
VERSION AJ414157.1
KEYWORDS Versinia pestis.
SOURCE Versinia pestis.
ORGANISM Versinia pestis
```

```
Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
Yersinia.
1 (bases 1 to 216050)
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
Fellwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,
Karlyshev,A.V., Moule,S., Oyston,P.C.F., Ouail,M., Rutherford,K.,
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Bartell,B.G.
Genome sequence of Yersinia pestis, the causative agent of plague
Nature 413 (6855), 523-527 (2001)
216050
2 (bases 1 to 216050)
Parkhill,J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/Y_pestis/).
Location/Qualifiers
1. 216050
/organism="Yersinia pestis"
/strain="CO92"
/db_xref="taxon:632"
/note="biovar: Orientalis"
86. 940
/gene="YPO3402"
/note="panC"
86. 940
/gene="YPO3402"
/note="Similar to Escherichia coli pantoate--beta-alanine
ligase PanC SW:PANC_ECOLI (P31663) (283 aa) fasta scores:
E(): 0, 72.0% id in 282 aa, and to Vibrio cholerae
pantoate--beta-alanine ligase Vc0591 TR:Q9KUD1
(EMBL:AE004144) (293 aa) fasta scores: E(): 0, 59.0% id in
283 aa"
/transl_table=11
/product="pantoate--beta-alanine ligase"
/protein_id="CAC92632.1"
/db_xref="GI:15981329"
/db_xref="SPTREMBL:O82BK7"
/translation="MLITETPLRQQIRRWQEGKRIALVPTMGNLHGHMTLVDEA
KTRADVVTIPVNPLOFERPDLAHYPTLOEDCEKLTGHGADLVFAPAAADIYPAG
LEKQYVDVPALSTILEGASRPGHFGYSTIVSKLFNLIQPDVACGEKDYQDLALIR
KVAWDGIDINIVGPTVRKDGALSSRNGLYETEEROIARPOLSKIMWALAEKMWALG
EROIDALLEAAAAQLLRVGFPTDELFIQDAETLQPLTVDSQQAVILMAWLKARLID
NQLVDLRH"
1351. 1731
/gene="YPO3403"
/note="panD"
1351. 1731
/gene="YPO3403"
/note="Similar to Escherichia coli aspartate
1-decarboxylase precursor PanD SW:PAND_ECOLI (P31664) (126
aa) fasta scores: E(): 0, 78.6% id in 126 aa, and to
Pseudomonas aeruginosa aspartate 1-decarboxylase precursor
Pa4731 TR:AAG08117 (EMBL:AE004887) (126 aa) fasta scores:
E(): 1.5e-27, 58.2% id in 122 aa"
/codon_start=1
/transl_table=11
/product="aspartate 1-decarboxylase precursor"
/protein_id="CAC92633.1"
/db_xref="GI:15981330"
/db_xref="SPTREMBL:O82BK6"
/translation="MRTMLQGLKLRVKVTQADLHYEGSCAIDQDFLEAAGILEYEA
```



```
QY 76 ThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAla 95
Db 136834 ----- 136834

QY 96 ProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu 115
Db 136833 ---GAAGCTATTGGGAGGGGCTGCGCGGATTAGGTTATTACATCCACCATAGCTTT 136777

QY 116 ThrGluLysAsp-----GlyAlaTyrThrValHisIleThrProGly 129
Db 136776 GAATTACAAACCGCCCGGACCGGACGCTTCAGTGTATTTCCAAAGAGTGGTCTCTGGC 136717

QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
Db 136716 GAGCCGGTATTGATTCGCGGTGTCGATCTACTCAAGCGGTGCGAAACGGATCCT 136657

QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAsp 169
Db 136656 GAT-----TATCAGGCACCTCTGTCGCCGTGATACCCCGAAATTTGGCTCGGTC 136609

QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
Db 136608 CTCACCATCGGTGATTGATAATTTCACGAGTCTCTTACTGGTTAGCGTTACGCCGT 136549

QY 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr 209
Db 136548 GGCTATTTCGACGCCAATATGATCAAAAGCCACGCTCGGTGTTCCCGCTCAATTACACGAG 136489

QY 210 ValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIle 229
Db 136488 GCGTTCGGGATATTGATTTCGACAGTGGGCAACGTTACCGCTTGTGAAGTTATTTC 136429

QY 230 ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGly 249
Db 136428 CAAGGCTCACAA---ATCCGAGAGGATTATCTGCAGAAATTAGTCGCGTTCCATGAGGG 136372

QY 250 ThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHis 269
Db 136371 GAATTCATACCTCTGATGAATGGCGGAGTTGCAACCGCGCTGGCAGCCACCACTGG 136312

QY 270 TyrSerGlyAlaSerValGlnAlaAspPhe---AspArgLeuGlnGlyAspArgValPro 288
Db 136311 TTTAACTCGGTGTTGTTTTCGCGGATTTCCAAAGATGCCAAAGAGAGTAAATTTTACCG 136252

QY 289 ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp 308
Db 136251 CTTGATCGCGTCTGTTACCCCA-----CGCACTGAG 136222

QY 309 SerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGly 328
Db 136221 AATACCTCGAACTGGGGGCGGTTATGCAACCGAT----- 136186

QY 329 TyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIle 348
Db 136185 ---GTCGGT-----CCGCGTCTGACCGCAGTTGG 136159

QY 349 SerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 136158 CGTAAGCCT-----TGGATGAACCTCTTCGGGCATAGCTTAACC 136120

QY 369 ThrThrGlnAsnLeu-----GluLysArgAlaPheSerGlyGlyIleTrpTyr 384
Db 136119 ACCACTACCGCGCTTCGCGACCTGACACACACTGGATTTCAGTACCGTATCTCTTTA 136060

QY 385 ValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly--- 403
Db 136059 TTGAGAAC-----CCCTCGAGCAGTATTACCTGATACAAGGGGC 136018

QY 404 ---ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet----- 420
Db 136017 TTTAGACGT-----ACCGACTTAACGACACTAATTCAGATACCCAGCCTTGAAC 135967
```

```
QY 421 -----LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
Db 135966 GTTGCACGTTTTTGGGATCTCTCCACGGCTGCAAAAGGCGGATT-----AACCTG 135916

QY 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
Db 135915 CGTTGGAGCCTCGATCACTTTACCAAGCAGAGTCACCGATACCATGTTGCTGTAT 135856

QY 455 SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr----- 468
Db 135855 CCGGGGCTGAGTATTATCTACCCGCCAACGTTGGTGGCGGATGCGGCTCTGGGGGAT 135796

QY 468 ----- 468
Db 135795 AGCCACGCGTATTCTATTGATGTCTGATACGACTTGGGGTGGATGTTGATTTTGGC 135736

QY 469 PhePheThrProGluAsn-----LysLysLeuGly-----ThrPheIleIle 482
Db 135735 ATCTTTTCAGCGCGCAGAACGTTGGATCCGACCTTAGGTGAGAAATCGTTTGTGGCG 135676

QY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
Db 135675 CGTGGCAATGTGGGGTGGATCGCAACCAATAATTTTACCGTGTTCACCGTCGTCGCT 135616

QY 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 135615 TTCTTCGCGGGGCGCAGCCAGTATCCGTTGGCTATAAATTCGTAATATTTCACCGCGT 135556

QY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
Db 135555 GATAGTGCAGCGAAATTAACCGGGGCTTCCAAAGTTGGCAACGGGCTCTCTTGAATACAA 135496

QY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 562
Db 135495 TATACTTCACGGTACAGTGGGGCGGCTTTTCGCTATTCGGCGAGGCGGTGAAT 135436

QY 563 AsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
Db 135435 AATTCAGCAAGAGTGACTTAAACACAGGGGAGGTGTCGGGTACGTTGGGCATCACCC 135376

QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys-----Ile 599
Db 135375 GTTGGACCGATCAAACTGGATATCGCGGCACCAATAGCGGATACGAGACACATGGCGTG 135316

QY 600 ArgTrpHisIleSerLeuGly 606
Db 135315 CAATTTCATCGGTTTGGG 135295

RESULT 18
LOCUS AE008906 24892 bp DNA linear BCT 31-JUL-2002
DEFINITION Salmonella typhimurium LT2, section 210 of 220 of the complete genome.
ACCESSION AE008906 AB006468
VERSION AE008906.1 GI:16422950
SOURCE Salmonella typhimurium LT2.
ORGANISM Salmonella typhimurium LT2.
REFERENCE 1 (bases 1 to 24892)
AUTHORS McLelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
Grewal,N., Mulvaney,E., Ryan,H., Sun,H., Florea,L., Miller,W.,
Stonking,T., Nhan,M., Waterston,R. and Willson,R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 24892)
AUTHORS The Salmonella typhimurium Genome Sequencing Project.
```

TITLE JOURNAL	Direct Submission Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA	CDS	841..1236 /gene="rpsf" /note="similar to E. coli 30S ribosomal subunit protein S6 (AAC77157.1); Blastp hit to AAC77157.1 (131 aa), 99% identity in aa 1 - 131"
COMMENT	COMMENT Supported by NIH grant 5U 01 AI43283 Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/ EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/ , and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/ The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.	gene	/transl_table=11 /product="30S ribosomal subunit protein S6" /protein_id="AAL23211.1" /db_xref="GI:16422953" /translation="MRHYEIVFMVHPDQSEQVPGMIERYSAATGAGKIHRLDWGR ROLAYPINKLKAHYVLMVPAEQEVIDLETTTFENDAVIRSMVMTKHAIVTEASPM VKADERRERDDFANETADDAAGDSEE" 1226..1557 /gene="priB" /note="STM4392" 1226..1231 /gene="priB" /note="putative RBS for priB; RegulonDB:STMS1H004277" 1243..1557 /gene="priB" /note="Similar to E. coli primosomal replication protein N (AAC77158.1); Blastp hit to AAC77158.1 (104 aa), 99% identity in aa 1 - 104" /codon_start=1 /transl_table=11 /product="primosomal replication protein N" /protein_id="AAL23212.1" /db_xref="GI:16422954" /translation="MTNRLALSGTVCRAPLRKVPSPGIPHCQFVLHRSVQBEAGFHR QAWCMPVIVSGHENOAITHSITVGSRTVQGFISCHKAKNGLSKMLHAEQELTDS GD" 1547..1789 /gene="rpsR" /note="STM4393" 1547..1552 /gene="rpsR" /note="putative RBS for rpsR; RegulonDB:STMS1H004278" 1562..1789 /gene="rpsR" /note="similar to E. coli 30S ribosomal subunit protein S18 (AAC77159.1); Blastp hit to AAC77159.1 (75 aa), 98% identity in aa 1 - 75" /codon_start=1 /transl_table=11 /product="30S ribosomal subunit protein S18" /protein_id="AAL23213.1" /db_xref="GI:16422955" /translation="MARYFRRRKFCRFTAGVQVEIDYKDIALTKNYITESGKIYPSRI TGTRAKYQRLRAIKRARIKRLPYTDRIQ" 1819..2280 /gene="rplI" /note="STM4394" 1819..1824 /gene="rplI" /note="putative RBS for rplI; RegulonDB:STMS1H004279" 1831..2280 /gene="rplI" /note="similar to E. coli 50S ribosomal subunit protein L9 (AAC77160.1); Blastp hit to AAC77160.1 (149 aa), 95% identity in aa 1 - 149" /codon_start=1 /transl_table=11 /product="50S ribosomal subunit protein L9" /protein_id="AAL23214.1" /db_xref="GI:16422956" /translation="MQVILLDKVANLGSQDQNVNKAAGYARNFVLPKGAIVPATKKNV EYFARRAELEAKLADVLAANAARAEKINALETVTIAKAGDEGKLFSGIGTRDIADA VTAAGVDVAKSEVRLPNGVLRITTTGHEVNFQVHSEVFAKVIINVAE" 2417..3354 /gene="yifZ" /note="STM4395" 2417..2422
FEATURES source	1..24892 /organism="Salmonella typhimurium LT2" /strain="LT2; SGSC 1412; ATCC 700720" /db_xref="ATCC:700720" /db_xref="taxon:99287" /note="LT2" complement(95..383) /gene="yjfY" /note="STM4389" complement(95..370) /gene="yjfY" /note="similar to E. coli orf, hypothetical protein (AAC77156.1); Blastp hit to AAC77156.1 (91 aa), 71% identity in aa 1 - 91" /codon_start=1 /transl_table=11 /product="putative outer membrane protein" /protein_id="AAL23209.1" /db_xref="GI:16422951" /translation="MRARIMLFUALLPGITATAAVALNNHQARNMDDVRSGLVIYIN HNPATSEANLALNEEADYRNAMYYHVIIRPGSNGNIHASANIYR" complement(378..383) /gene="yjfY" /note="putative RBS for yjfY; RegulonDB:STMS1H004274" 560..770 /gene="STM4390" 560..565 /gene="STM4390" /note="putative RBS for STM4390; RegulonDB:STMS1H004275" 564..770 /gene="STM4390" /note="hypothetical protein" /codon_start=1 /transl_table=11 /product="putative cytoplasmic protein" /protein_id="AAL23210.1" /db_xref="GI:16422952" /translation="MWQKSALTTSISLPQQAVYSCMRVEVNSAAKCDGTGVTFTGI NNTITLSLPAGASIIIRHFSADL" 828..1236 /gene="rpsF" /note="STM4391" 828..833 /gene="rpsF" /note="putative RBS for rpsF; RegulonDB:STMS1H004276"	gene	
		RBS	
		CDS	
		gene	
		RBS	
		CDS	
		gene	
		RBS	

Qy	208	aThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheG1	228
Db	16904	TCAGCGCTTTTGGGATATGATATGATAGCGGCTATCGCTTTCAGCACCTGTCCAC	16963
Qy	228	uIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaAspPheGlnPr	248
Db	16964	CTTCGAAGTTCGCAG---ATTCTGTGATGATATTACAAATCTGTTCCTGTTAAAG	17020
Qy	248	oGlyThrProTyrAspLeuAspPheGlnGlnAlaLeuGluGlnAsnG1	268
Db	17021	GGTGCATCAGTACGAATCGAAAGATTTCGGCGAATAAACCAGCGACTTTCGCTACCGG	17080
Qy	268	yHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVal--	287
Db	17081	ATGGTTTAAATCTGTCTGGTGGCCCGGAATTTGAGAAATCGCTAAACAAAATCTCT	17140
Qy	288	-ProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLe	307
Db	17141	CCCGTTAAAGCGGTGGTATCGCCGCAACCGAAACACGATTGAAACCGGGTGGGTA	17200
Qy	307	uAspSerGlyThrGlyLeuGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLys	327
Db	17201	CTCCACCACGTCGGCGCGCGTGAAGCGCTGTGG-----AAAAA	17242
Qy	327	sGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaG1	347
Db	17243	ACCGTGGATGAACCTTAC-----GGTCATACCTGCACACCGACACCGAC	17287
Qy	347	yIleSerGlnProArgAsnTyrArgLysAsnTyrTrpThrSerAsnValSerTyrAsnAr	367
Db	17288	TATTCTCGCGGGAACAG-----GTGCTGGGATTTTGGTTATATAAAT	17329
Qy	367	gSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTyrTyrValArgAs	387
Db	17330	CGCGCTGCTGAAAC-----	17345
Qy	387	pArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly-----ArgLy	405
Db	17346	-----CCGCTCGCAACTACTATCTGTTACAGGCGGCTTTTAAACG	17386
Qy	405	sIleProGlySerAspIleAsp-----LeuGlyAsnSerHisAlaThrMe	420
Db	17387	TACCGATTGGAATGATACGGAGCAGGACTCGACACCGTTCGCGTCTACTGGGA	17448
Qy	420	tLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHi	440
Db	17447	CTTTCACAGTGTTCGACGGCGCATTT-----AATCTGCCTGGAGTTTCGACCA	17497
Qy	440	sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu1	460
Db	17498	CTTTACCCAGGGGAACGTCACCAATACCACCATGTGTTCTACCGGCGGTGATGATCAG	17557
Qy	460	eArgThrSerAlaArgAlaGly-----TyrPhePh	470
Db	17558	CCGAACCGCTCGCAGGGGGGACTGATGCCAACCTGGGGCGATTTCGACGCGCTTTCAGT	17617
Qy	470	eThrProGluAsnLysLysLeuGlyThr-----	479
Db	17618	AGACTATTGCAATACGGCTGGGCTCCGACGTGGATTTCCTCGTCTCAGCGCAAAA	17677
Qy	480	-----PheIleIleArgGlyGlnAlaGlyTy	488
Db	17678	CGTCTGGATTCCGACTTTGTACGATCGCACCGTTTTGTGATGCGCGCAATTTGGCTG	17737
Qy	488	rThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAl	508
Db	17738	GATGAACCCGCTGATTTCGACAAAGTTCGCGCGGATTTACGTTCTTCGCGGGGGCGA	17797
Qy	508	sSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVa	528
Db	17798	CCGCAAGTATTCGCGCTATAATACAAATCTATTTCCGCTAAAGATAGCAGCGCAATCT	17857

```

Qy 528 lLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgTh 548
:      |||      |||      |||      |||      |||      |||      |||
Db 17858 TAAAGCGCCCTCAAAACTGGCAACCGGATCGCTGGAGTACCAATAACGTCACCGGTAA 17917

Qy 548 rLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLy 568
      |||      |||      |||      |||      |||      |||      |||
Db 17918 ATGTGGGGGCGAGTGTGTTCGATACGGCGGAGCGGTGAGTCATATTCGCCCTAGCGA 17977

Qy 568 sLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPh 588
      |||      |||      |||      |||      |||      |||      |||
Db 17978 TTTCAAAACGGGACCGGGTCGCGCTGCGCTGCGCGCTGGCGCTGTCAAACT 18037

Qy 588 eAspIleAlaTyrGlyHisSerAspLysLys-----IleArgTrpHisIleSerLe 605
      |||      |||      |||      |||      |||      |||      |||
Db 18038 CGATTTTGGCGTACCGGTGGCGGACAAAGACGACACGGTTTACAGTTTATATACGGTCT 18097

Qy 605 uGly 606
      |||
Db 18098 GGG 18101

RESULT 19
AL627283
LOCUS
DEFINITION
Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome; segment 19/20.
ACCESSION
AL627283 AL513382
VERSION
AL627283.1 GI:16505370
KEYWORDS
Salmonella enterica subsp. enterica serovar Typhi.
SOURCE
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 249050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Felwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Kroh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrell,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)
JOURNAL
MEDLINE
21534947
PUBMED
11677608
REFERENCE
2 (bases 1 to 249050)
Parkhill,J.J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/s\_typhi/).
Location/Qualifiers
1..249050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
complement(78..347)
/gene="STV4596"
complement(78..347)
/gene="STV4596"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD06718.1"
/db_xref="GI:16505371"
/db_xref="SPTREMBL:O821H0"

gene
source

gene
CDS

translation="MRINVLVLLTSLVAGPALAGEAHVCKVSQTVANSANAELTNTV
FKGCSIGTIPSLAREGKIVQQTQADVDTPSKTYAQLIIQKD"
complement(479..1753)
/gene="STV4597"
/notes="samba"
complement(479..1753)
/gene="STV4597"
/notes="Similar to Salmonella typhimurium cryptic 60-MDA
plasmid Samba protein SW:SAMB_SALTY (P23832) (424 aa) fasta
scores: E(): 0, 89.6% id in 423 aa, and to Shigella
flexneri impB impB TR:092A91 (EMBL:AF079316) (423 aa)
fasta scores: E(): 0, 71.7% id in 421 aa
paralogue of E. coli umuc (UMUC_ECOLI); Fasta hit to
UMUC_ECOLI (422 aa), 63% identity in 420 aa overlap"
/codon_start=1
/transl_table=11
/product="UV protection protein"
/protein_id="CAD06719.1"
/db_xref="GI:16505372"
/db_xref="SPTREMBL:O821G9"
/translation="MFALADVNSFYASCEKVFRRDLRNTPVVLSNNDGCVIARKEA
KQICIKMGVPFOLKTEPPVPVAFSSNYELYSKSRVYMAHLEELAPRVEQYSIDE
MFUDVRNIDSCIDFEDFQRLREHVSGLTIGVGMGPKTKLAKSAQWASKEMRQFE
GYLALVSGNPKRTEKLSLPQVDEIMGVGRRIKKLHTMTGTTALQALARANPTFIRKN
FSVLTERTRELNGESCISLEEAAPPQQIIVCSRFGERVTYIEMRQAYCOYFIARAA
EKLGRGRCRHVAVFVKTSFPAVNPYVGNVASEKLTPTQDTRDIIAAAVKALDRI
WVKGHRVAKAGCMNDPTGTGISQNLNFDVQPERSEQLMKVLDSINHSGMKVWPA
GRGTAPEWQMKRELLSPATYTRWALPVARLR"
complement(695..1741)
/gene="STV4597"
/notes="Pfam match to entry PF00817 IMS, impB/mucB/samb
family, score 545.70, E-value 3.1e-160"
complement(join(1753..1830,35386..35730))
/gene="STV4598"
/notes="samba"
/pseudo
complement(join(1753..1830,35386..35730))
/gene="STV4598"
/EC_number="3.4.21.-"
/notes="Similar to Salmonella typhimurium Samba protein sama
SW:SAMA_SALTY (P23831) (140 aa) fasta scores: E(): 0,
82.1% id in 140 aa, and to Salmonella typhimurium impA
protein impA SW:IMPA_SALTY (P18641) (145 aa) fasta scores:
E(): 1.5e-31, 64.2% id in 123 aa. This CDS has been
interrupted by the insertion of a bacteriophage"
/pseudo
/codon_start=1
/transl_table=11
/product="UV protection protein (pseudogene)"
/db_xref="PSEUDO:CAD06720.1"
/db_xref="RENTREMBL:CAD06720"
1831..1839
/notes="phage insertion point. Note this sequence is
duplicated at position 4507385..4507393"
1840..35385
/notes="bacteriophage"
1973..2350
/gene="STV4599"
1973..2350
/gene="STV4599"
/notes="no significant database hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD06721.1"
/db_xref="GI:16505373"
/db_xref="SPTREMBL:O821G8"
/translation="MNKPLIALLLITFSASADKIPSSIQNLIAVDTTRTHSLDGL
TIRYKNRLLVDAESMFGICNDYMNKWKPKTIKRITLLNVTSDQGFPEINAGDDC
RKAGTKWDEQARTYTSFIKPLQ"
complement(2437..2655)
/gene="STV4600"
complement(2437..2655)

```



```

/ gene="STY4600"
/ note="Similar to Escherichia coli prophage P2 Ogr protein
SW:OGRK_ECOLI () (72 aa) fasta scores: E(): 4.8e-15, 55.7%
id in 70 aa and to Bacteriophage 186 late gene control
protein B SW:VPB_BP186 (P08711) (72 aa) fasta scores: E():
2.1e-17, 64.2% id in 67 aa."
/ codon_start=1
/ transi_table=11
/ product="putative positive regulator of late gene
transcription"
/ protein_id="CAD06722.1"
/ db_xref="GI:16505374"
/ translation="MMNCPKCGHSARTRSSFQVTDSTKERYCQCQINCGSTFVTHET
VYRFIVTALVNNAPHPHTVSGQHMF"
/ complement(2723..3823)
/ gene="STY4601"
/ complement(2723..3823)
/ note="Similar to Bacteriophage P2 late gene control
protein D SW:VPD_BPP2 (P10312) (387 aa) fasta scores: E():
0.54.3% id in 370 aa and to Bacteriophage 186 late gene
control protein D SW:VPD_BP186 (P21679) (389 aa) fasta
scores: E(): 0, 54.7% id in 371 aa"
/ codon_start=1
/ transi_table=11
/ product="putative regulator of late gene expression"
/ protein_id="CAD06723.1"
/ db_xref="GI:16505375"
/ translation="MNFSSDLFLGNSRSPAFSITIEGKDVTTALDARLMSLTLDNRG
FEADQDLDDADGQIVLPRGAVIOLALGWKGPFPKGAFTVDEIHSAGPDRLT
IRASADPRLTIRREKSHOTTVEGVKETAARHNLKVALGKDLTKVDLHMDQTN
EDASPLKRLAQYGAISVKDGNLLFIROCGRTASCKPLPVIITTRKAGDHFRFL
ADGATGVIAISLITREPKKETTQVRHRKAVATPEKPEAKQGDYLVKGTDENLVL
WNTITLHTVSPDNGFTTSLELVKIDDLIE"
/ complement(3820..4305)
/ gene="STY4602"
/ complement(3820..4305)
/ note="Similar to Bacteriophage P2 complete genome U
essential tail protein TR:O64315 (EMBL:AF063097) (159 aa)
fasta scores: E(): 2.2e-26, 45.2% id in 157 aa and
Bacteriophage 186 protein F TR:O80317 (EMBL:U32222) (161
aa) fasta scores: E(): 1.1e-29, 49.1% id in 159 aa."
/ codon_start=1
/ transi_table=11
/ product="putative phage tail protein"
/ protein_id="CAD06724.1"
/ db_xref="GI:16505376"
/ translation="MMVLGLYFVHLRTVPYQELQYQSRHUAANSRVNRRSTQFLG
PNDPJTLSGLVLPVETGGRLSLLALEQMAELGKAWPLLEGSGTIGYMFVIESLSQTK
TEFFSGMRRRIETLTLKRVDIESLDMFGSLSDLSLQNSQDSATSAIGNIKNTVGGLL
Q"
/ complement(4305..7085)
/ gene="STY4603"
/ complement(4305..7085)
/ note="Similar to Bacteriophage 186 protein G TR:Q37848
(EMBL:U32222) (812 aa) fasta scores: E(): 1.4e-14, 24.5%
id in 694 aa"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="CAD06725.1"
/ db_xref="GI:16505377"
/ translation="MSDNNLRLOVVLGAVDKLTPFKNAQASKELASAIRQTRDRIK
KUSDAGQKSLDQLTQSVSRGAEIDQARLRAOMMTREMSLESPTKKOTQALEAQW
RAVSRLEQKQOETROMAAARELYRLGLSAGGARETARIARETERYNRLAEQERR
LREVGSRORKLNAIKAKAEKTRLNRLNLAGNGAMAAGVTTGMTLLAPVKVASEN

```

```

Alignment Scores:
Pred. No.: 2,22e-12 Length: 249050
Score: 357.00 Matches: 139
Percent Similarity: 39.53% Conservative: 99
Best Local Similarity: 23.09% Mismatches: 262
Query Match: 11.25% Indels: 103
DB: Gaps: 18

US-09-857-669-2 (1-609) x AL627283 (1-249050)
QY 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
Db 153234 CTGAAGTCGAAGGGCTATCCGGAGAGCTGGAAGAAACGTTCCGGCACACGCTCTCTACG 153293
QY 75 IleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGlu 94
Db 153294 ATT-----CAGAGCGATGAGGTGACGCGGATCGGCGC-----TTTCGCGCCCGA 153338
QY 95 AlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 153339 GTTGATGATCGGATTCGGAAGGGCTTAAAGCGTTAGGCTATTACGAACCCACTATCAA 153398
QY 115 Leu-Thr-----GluLysAspGlyAlaTyrThrVal---HisIleThrPr 128
Db 153399 TTCGACTTGGTTCACCGCTCGGAAGAGCGGAGGTATTA-ATGCCACAGGGTTAGCC 153457
QY 128 OGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAs 148
Db 153458 CGCCACCGCGGTGTTGATTCGGGTACAGAAAGTCATTTACGCGCGCGCGCTACCGA 153517
QY 148 pGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySe 168
Db 153518 C-----AAAGATTATCTG-----GCCTTACTGAAACGCGTCCGCAATCGGCAC 153562
QY 168 rAspPheAspGluAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrAr 188
Db 153563 GTTGCTCACCGAGGGGACTATGACAAATTTAAAGCTGTTAACAGCTGTCGCTGCTACG 153622
QY 188 gLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAlaAsnProAspThrAl 208
Db 153623 CAAAGGCTATTTCGACAGCGAATTCATTAAGTCAGTTCGGAAATTCCTCTGCGCCGTCA 153682
QY 208 aThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGl 228
Db 153683 TCAGCGCTTTTGGGATATTGATTATGATAGCGGTGACGCTATCGCTTCGCGCTGTCAC 153742
QY 228 uIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnPr 248
Db 153743 CTTCGAAGGTTCCGAG---ATTGCTGATGAATATTTACAAATCTGTTCGGCTTTAAGA 153799
QY 248 OGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGl 268
Db 153800 GGGTGATGACTACGATCGAAGATTTGGCGGAACCTAAACCGCGCGCTTCCGCTACCG 153859
QY 268 yHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVal-- 287
Db 153860 ATGGTTTAAATCTCTTGTGGCGCGGAAATTTGACAAATTCGGGATAAACAATAATCT 153919
QY 288 -ProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLe 307
Db 153920 GCCGTTAAAGGGGTGGTATATCACCAGCAACCAACACCATGTAACCGGGTGGGGTA 153979
QY 307 uAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys 327
Db 153980 CTCCACCGAGCTCGGGCGCGCGGTGAAGCGTCTCTGG-----AAAA 154021
QY 327 sGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaLagL 347
Db 154022 ACCGTGATCAACTCTCTAC-----GGTCATAGCTGTGACCAACCAACCAACCAAC 154066

```

```
Qy 347 yileSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnAr 367
||||| |||
Db 154067 TATTCTCGCCGGAACAG-----GTCTGGATTTTACTTATAAAT 154108
|||
Qy 367 gSerThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAs 387
|||||
Db 154109 CCGCTCTGAAAAC----- 154124
|||
Qy 387 pArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly-----ArgLy 405
||||| |||
Db 154125 -----CCGCTGGAGCAATACTATCTGTACAGGCGGTTTAAACG 154165
|||
Qy 405 sileProGlySerAspIleAsp-----LeuGlyAsnSerHisAlaThrMe 420
||| |||
Db 154166 TACGATTTCGATGATCAGGAGCAGGACTCGACCACCGCTTCCGCTGCTACTCGGA 154225
|||
Qy 420 tLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHI 440
||||| |||
Db 154226 CCTTTCAGCGGTTGGCAGCGGCGGATT-----AACCTGCGCTGGAGTTTCGACCA 154276
|||
Qy 440 sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuII 460
||| |||
Db 154277 CTTTACCCAGGGGAACGTACCAATACCAATACCATCGCTGTTTACCGCGCGTGATGATCAG 154336
|||
Qy 460 eArgThrSerAlaArgAlaGly-----TyrPhePh 470
||||| |||
Db 154337 CCGAACCCGCTCGGAGGGGACTGATGCCAACCTGGGCGGATTCGCGAGCGCTACTCGT 154396
|||
Qy 470 eThrProGluAsnLysLysLeuGlyThr----- 479
|||
Db 154397 AGACTATTTCGAATACGCGCTCGGCTCCGACGTGATTTCTCCGTACTCGAGCGGCAAAA 154456
|||
Qy 480 -----PheIleArgGlyGlnAlaGlyTy 488
||||| |||
Db 154457 CGTCTGATTTCGGAGCTTTGTACGATCCGACCGCTTTGTGATGCGCGCCAAATTTGGGCTG 154516
|||
Qy 488 rThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAl 508
||||| |||
Db 154517 GATAGAACCGGTGATTCGACAAAGTTCGCGGATTTACGTTTCTTCGCGGGGCGGA 154576
|||
Qy 508 aSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVa 528
||||| |||
Db 154577 CCGCAGTATTTCGGCGCTATAACAAATCTATTTCGCTTAAAGATAGCAGCGCAATCT 154636
|||
Qy 528 lLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArgTh 548
||||| |||
Db 154637 TAAAGCGCCCTCAAACTGGCAACCGGATCGCTGGAGTACAGTATTAACGTCACCGCAA 154696
|||
Qy 548 rLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhelYsArgMetLy 568
||||| |||
Db 154697 ATGCTGGGGCGGTGTTTCGATACGCGGAGCGGTTAGTATTCGCGGTAGCGA 154756
|||
Qy 568 sLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPh 588
||||| |||
Db 154757 TTTCAAAACCGGACCGGCTCGCGTTCGCGTTCGCGCTCGCGCTGTCGCAACT 154816
|||
Qy 588 eAspIleAlaTyrGlyHisSerAspLysLys-----IleArgTrpHisIleSerLe 605
||||| |||
Db 154817 CGATTTTCGCTACCGCTCGCGCAAGAGAACGACGCGTTTACAGTTTATATPCGCTCT 154876
|||
Qy 605 uGly 606
|||
Db 154877 GGGG 154880
```

RESULT 20

```
AX033470
LOCUS
DEFINITION Sequence 3 from Patent WO0047737.
ACCESSION AX033470
VERSION AX033470.1 GI:10280230
KEYWORDS Haemophilus influenzae.
SOURCE
```

```
ORGANISM Haemophilus influenzae
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 1731)
AUTHORS Ruelle,J.L. and Thonnard,J.
TITLE Haemophilus influenzae rd outer membrane sequences used as vaccine
JOURNAL Patent: WO 0047737-A 3 17-AUG-2000;
RUELE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ; THONNARD
JOELLE (BE)
FEATURES
source Location/Qualifiers
1..1731
/organism="Haemophilus influenzae"
/db_xref="taxon:727"
BASE COUNT 509 a 322 c 386 g 514 t
ORIGIN
```

Alignment Scores:

```
Pred. No.: 4,13e-15 Length: 1731
Score: 355.50 Matches: 138
Percent Similarity: 38.01% Conservative: 87
Best Local Similarity: 23.31% Mismatches: 238
Query Match: 11.20% Indels: 129
DB: 6 Gaps: 15
```

US-09-857-669-2 (1-609) x AX033470 (1-1731)

```
Qy 79 GlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsn 98
||||| |||
Db 145 AAGAAGAAATGAGCGGCTCTGAACGCTATCAGCATTTGGTAACCAANGCGGTGATCGT 204
|||
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerLysValSerLysLeuThrGluLys 118
||||| |||
Db 205 -----GGTTGCGGTGTTTGGTTATTATGAATCTTCCGTGCGTTTGAACGAAA 255
|||
Qy 119 Asp-----GlyAlaTyrThrValHisIleThrProGlyProArgThrLys 133
||||| |||
Db 256 CAGCGTCAAGGCAAAACCGGATTTATTGATTCGCTCATCCAGCGGAGCCAAACAAA 315
|||
Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu 153
||||| |||
Db 316 ATTCGGGGGACTGATGTGCANATTGAGGGGAGAGCCCAACAGATGAAATTTT- 369
|||
Qy 154 TyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAsp 173
||||| |||
Db 370 -----GATGCGCTACGTAAAAACTTGCACAAAGAGCGGTTTGGTTGAACACCAA 420
|||
Qy 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu 193
||||| |||
Db 421 ACTTACGATGATTACAAAACAGCGGATTTTCACGCTTGCATTAATCGTGGGTATTTTGAT 480
|||
Qy 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
||||| |||
Db 481 GGGNACTTTAAATTTCCAGCTTTAGAAATCAGCCCTGAAACCCATCAAGCATGTGGCGGA 540
|||
Qy 214 ValValAspSerGlyArgProIleAlaPheGlyPheGlyIleThrGlyThrGln 233
||||| |||
Db 541 ATGTTATTGATAGTGTGCGTTATCATTTATGTCATATATTTACTTTTAGCCATTACAA 600
|||
Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
||||| |||
Db 601 ---ATCCGTGACGATTTATCTCAATATATTTCTTAACATCAAACTGCGCGATTTTAA 657
|||
Qy 254 LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
||||| |||
Db 658 ATGATTAATTTGTCGGATTTAACCCAGCGATTTTTCATCTTCAATTTGGTTAGCTCAGTA 717
|||
Qy 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
||||| |||
Db 718 TTAGTTTCCAGCTAATGTTAATCAT---AAAAGCAAAACTGTGGATGTGGAGATTATTTCTT 774
|||
Qy 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
||||| |||
Db 775 TATCCACGTAAAAAAATGCGATGGAACGCGTGTGCTGCTGCGCTTGTCTACTGCGGCGGT 834
|||
```

```

RESULT 21
AE004323/c
LOCUS
DEFINITION
14390 bp DNA linear BCT 31-JUL-2000
Vibrio cholerae chromosome I, section 231 of 251 of the complete
chromosome.
ACCESSION
AE004323 AE003852
VERSION
AE004323.1 GI:9657119
KEYWORDS
SOURCE
Vibrio cholerae.
ORGANISM
Vibrio cholerae.
REFERENCE
1 (bases 1 to 14390)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Seillers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
TITLE
Nature 406 (6795), 477-483 (2000)
JOURNAL
MEDLINE
20406833
PUBMED
10952301
REFERENCE
2 (bases 1 to 14390)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Seillers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
TITLE
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
JOURNAL
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1. 14390
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="biotype: El Tor"
80. .661
/gene="VC2536"
80. .661
/gene="VC2536"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95677.1"
/db_xref="GI:9657120"
/translation="MKMKARKNQKAPWEEEEI IWSRTELKNDMLALQKLGELVE
LKPSALAKFPLPEDLAETADQRFKNEARRRQLOVIGKLMRHIDPEPLQAAIDKLRLN
KHSSTALLHLKROLRIVAEGDSAIEVAMEQYPEADQRQLRLARQASKERAGNKP
PKSREIFOLLKEAMLAQKQIEEEDLDSAE"
complement(764. .1498)
/gene="VC2537"
complement(764. .1498)
/note="VC2537"
/note="similar to GB:D10483 SP:P31548 PID:216486 GB:U00096
PID:1786253; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="thiamine ABC transporter, ATP-binding protein,
putative"
/protein_id="AAF95678.1"
/db_xref="GI:9657121"
/translation="MLMRKSNLALDKVREYEHWFEDLNVDGDIVALMGPSGAGK
STLLSLVAGTIEPVSGSIKVDQSVGLAPYQHPFSLFOEHLFAHLYVRENIIGL
HPGLKLNAGKQVDAQQVGIADYLDRLPEQLSGGQRQVRLARCFVQPNPIWLD
EPFSALDPLRLREMLALVKQLASERQRTVVMVTHLSLSDARAIASQIAFLSQGKVKVS
DCQAVTAQHHPHPELAQFVAALNEPK"

```



```
Db 14093 CGTTTCAATCTCAACTGGAAGAAGATGACC-----GAAGCG----- 14055
QY 98 AsnValLysThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeuThr--- 116
Db 14054 -----TTGAATGCACTCGGCTACTACCCAGTATCGATTTTACTGTGC 14010
QY 117 ---GluLysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAla 135
Db 14009 TCTGAAGATAACACGCGTTTACGCGCGGAGTGACCTTGGCGAAGTACCGCTTGAGC 13950
QY 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspClyAsnLeuAlaGluTyrTyr 155
Db 13949 GAGGTTGATATTGTATCCGCGTGAAGCGGAGCGGATGCGGATTTTCAACGCTTTGATC 13890
QY 156 ArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrp 175
Db 13889 CGCGCGCAGT-----GCTTTGAGAGTGGATGCGCGCTTAAATCACAGTCTGTAT 13842
QY 176 GluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLys 195
Db 13841 GACAACCTCAAAATCGGCGATTCGTAATCGGCTTGAAGGCTTACTTTAATGGTGAT 13782
QY 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
Db 13781 TTTCAGCGCCAGCGCGCTGGAAGTGTATCCAGAGCTGAATCAAGCTCGGCTGATTTTGCAC 13722
QY 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyr 235
Db 13721 TTTGATAGCGGATCGCTTACCTGTTTGGAGCGACCGCTCAAGCGCAGTCAA---ATC 13665
QY 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 13664 GATCAAAATCGCGTGATGTCGCTGCGCTTTCACAAAGCGGAGCCTTATTTGGTGTGC 13605
QY 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerVal 275
Db 13604 CAATCGCGGAGTTTAAACCAAAATCTTCCAATACAGATGGTCTCTCGGTGTTTGT 13545
QY 276 GlnAlaAspPheAspArgLeu---GlnGlyAspArgValProValLysValSerValThr 294
Db 13544 GACCGCGACTTATCGCAGTTGGATGAGGCGGTGAGTTCGGGATCAAGGTGACCTTAGCA 13485
QY 295 GluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGly 314
Db 13484 CCGCAAGCAGCGCAATCAACTGGAACCGGTTTAGGCTATTCCACGCGGTTGGAGTAGCT 13425
QY 315 GlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValVal 334
Db 13424 GGTTCATTAAATGGAAGAACCTTGGGTGAACAGCGCAAGGGCAGCATTTTGACAGCAGT 13365
QY 335 TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro----- 351
Db 13364 TTCTCTCTTCTATACAGAGCAGACCATTTACGCGGGTTACAGATCCCACTCGGAAGAT 13305
QY 352 -----ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 13304 CGGTTAAATCAGTACTATCGCATTCAGTACGGGATGAAGCATTTG-----GACAAGCGT 13251
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyClyIleTrpTyrValArgAspArg 388
Db 13250 GATACCGAGAGCTTGAATCTAATCTATCATCTTGAACGCCATTGG----- 13206
QY 389 AlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluArgLysIleProGly 408
Db 13206 ----- 13206
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGln 428
Db 13205 ---CAGTTGGAT---GGCGGCTGGCATCGCAGCGTCTTTATT-----CGTTAT 13164
QY 429 LeuLeuAsnAsnVal-----LeuHisProGluAsnGlyHisTyr---LeuAspGly 444
Db 13163 CTCTTAGAGAATTATCGTCAGGGTTTACAAAGATGATAACAGCAATTTTATTTCGCGGGG 13104
```

```
QY 445 LysIleGlyThrThrLeuGlyThr----- 452
Db 13103 ATGACTTATACCGTCGTCGCGGACTCGCTCCAACAGTGGCTTACTGACTTGGGGCGATAA 13044
QY 453 -----PheLeuSerSerThrAlaLeuIleArg 461
Db 13043 CAGACGATTACCTTGGAAATATGCGATCCAGCATTCGTTTTCAGAAACACGGGTATTGCGT 12984
QY 462 ThrSerAlaArgAlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThrPhe 480
Db 12983 TTACAACCGGATCTCTTCTGGTTACGAATTTATGCGGCAANTCACCAGCGCC----- 12933
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 12932 TTAGTTTCGCTGATGGTGGAGCCAATTTGGTGGATGAATTTCCACCAACTCTCGCCCTCG 12873
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 12872 CTGCGTTTCTTTTCCGCTGGTGATACAAATTTTCGCGGATACGTTTACAAGTCGATTCA 12813
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGlu 540
Db 12812 CCCACAGATGCCAGTGGCGCTTTTGACAGGGCGGAATATACATCCGACCAGCTCAATCGAA 12753
QY 541 TyrGlnLeuProPheThrArgThrLeuSerGly-AlaValPheHisAspMetGlyAspAl 560
Db 12752 TATCATA-CCGTTTAAACCGCAACTGGTGGCGCGCATGTTTCATGATGTCGCGTATGC 12694
QY 560 aAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPh 580
Db 12693 GTTAAACGAT-----AATCCTCAATGGAAGAAAGGGTCGGTACTGGGATCGCTGGAT 12640
QY 580 eSerProLeuAlaProPheSerPheAspIleAlaTyrGly-----HisSerAspLy 597
Db 12639 TTCGCGCGTTGGCGCTATTTCGGCTCGATTTTCGATGGGGCTGGATGCCGCGCGGCGA 12580
QY 597 sLysIleArgTrpHisIleSerLeuGly 606
Db 12579 TGAGTTCAAAATCCACTTCACTTTAGGG 12552
RESULT 22
AL646064 201050 bp DNA linear BCT 07-DEC-2001
LOCUS Ralstonia solanacearum GMI1000 chromosome, complete sequence;
DEFINITION segment 8/19.
ACCESSION AL646064
VERSION AL646064.1 GI:17428340
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum.
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 201050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Benard,C., Cunnac,S., Denance,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schlex,T.,
Sigulier,P., Thiebaut,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 201050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
```

F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES

source

1..201050

/organism="Ralstonia solanacearum"

/strain="GMI1000"

/db_xref="taxon:305"

206..1447

/gene="RSC1326"

/note="RS02848"

206..1447

/gene="RSC1326"

/EC_number="2.6.1.1"

/function="small molecule metabolism; amino acid

biosynthesis"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABE ASPARTATE AMINOTRANSFERASE PROTEIN"

/protein_id="CAD15028.1"

/db_xref="GI:17428341"

/translation="MKTQKSAKLNVCYDIRCPVLEKAKQMEESCHQIKLNIGNLA

PRGPAPETIQDMIRNLPNSAGYSDSKGIFAARKAWHYTOQOGIKNVLDIILGN

GASELIATNALNAGDELPLPADYPLWTAMTSLSGTTPVHYLCDESNWMPLODO

IRAKITPTKIGVIVNPNNTGALYSDALLHDIATAREHGLVIFADEVYKVLFDON

RITAMASLSEDEVLTFTNSLSKSYRSCGYRAGVMVYSGDKRPADKYIEGLNKLSSMRL

CVNQAIOALGQYOSINDLVAPGGMRRQRLAYELITAIQVSCVKPKAALYM

FRLPDSVPIDDDQTFIRQLLEERVLVQGTGFNWHSPDFRIVFLPHEDDLREAI

GRIARFLERYRHRGTGIRAA"

1538..2857

/gene="RSC1327"

/note="RS02849"

1538..2857

/gene="RSC1327"

/EC_number="1.1.1.1"

/function="small molecule metabolism; amino acid

biosynthesis; threonine"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABE HOMOSERINE DEHYDROGENASE OXIDOREDUCTASE

PROTEIN"

/protein_id="CAD15029.1"

/db_xref="GI:17428342"

/translation="MKPIKIGLGLTGGGTGLKVLORNOEIRRRAGRGIEVALIAV

RVERARATVQETGASVPVDDPVSPEIDIVVELIGGYDLTRRLVLRAIANGKH

VYTANKALLAHNGELFKAAQDKGVNAFAEAVAGGIPILKALREGLTGNRIEWTAGI

INGTNFILSEMRDKLQDGTVLAEARQIRYAEADPTFDIEGFDRAHKLTLLMSAIFG

VPVQERAVVEGITKLEAVDIIHAEELGYRIKLGLTRAEAGIELRVHTLVPATRL

IANVCAMNAVQGDVAGPTLYYKGGAEPATASAVADLVDTVRLHTADPEHRVPH

LAFOQDALNSLTPLLPEEVSAYILRMVADETGVLADITRILAESGSIIDAMLKQES

RGEPTQDTIIMLSHVVEKQVNAATAIEALPTVLSKVTRLRMEELN"

2889..4334

/gene="thrc"

/note="RSC1328; RS02850"

2889..4334

/gene="thrc"

/EC_number="4.2.99.2"

/function="small molecule metabolism; amino acid

biosynthesis; threonine"

/note="Product confidence : probable

Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABE THREONINE SYNTHASE PROTEIN"

/protein_id="CAD15030.1"

/db_xref="GI:17428343"

/translation="MQYRSTRGHAEPQSFRIILGGLAPDGGLYLPEQYPOVSABELT

RWRDLPYADLAFAEILRKFKATDIPEDDLRALTRRTTYPEVCNARNGDNTADITPLHRL

GWEAGTSLSLCLSNQPTLAFDMAQMLGNLFYALDRHAEILNITLGSDTGUSAA

EYAMRGKRGVRVMSIPHRKMSAFQTAQFSLQDKNIFNIAIEGVFDDAQDITKAVSN

DLEYKAROKIGTVNSINWARVAQVYVFKGYLLAAKPIGDKVSCVPSGNGFNVACG

HIARMNGLPIDKLVLVATNENDVLDDEFRTGTYRVSAAQTYHTSSPMDISKASNFER

FVFDLLGRDAPRLAQFRDVKDEKGGFSGMTGKPEFDRIABEFGVSGRSTHEDRVNTRD

VFVRYGTWVDTHTAGVAVARENLPVGVPMVVLLETALPAKFGDTIREALDREPERPAG

YADIESLPQRFVEMPADADRIRKGYTAQTGL"

4361..4957

/gene="RSC1329"

/note="RS02851"

4361..4957

/gene="RSC1329"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="CONSERVED HYPOTHETICAL PROTEIN"

/protein_id="CAD15031.1"

/db_xref="GI:17428344"

/translation="MAEAAFAFASCLHDIRACICEGALPHEPRPVVLADRARRVLIVGQ

APGRIVHETGIPWNRSGDLREWLRIQYDAFYADRRIAVVPMGFCFPGTGNGLDLP

RPACAPQWHRPLLTAMPDLKLALLVGYAQRYVLDVPRGMPLTEVVRGPGVDAEGRV

LFPLPSPRNLAWEKHPWFDAQIVPLRAALAAAFV"

5037..6296

/gene="moeA2"

/note="RSC1330; RS02852"

5037..6296

/gene="moeA2"

/function="small molecule metabolism; biosynthesis of

cofactors, carriers; molybdopterin"

/note="Product confidence : putative

Gene name confidence : putative

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE MOLYBDOPTERIN BIOSYNTHESIS MOBA

PROTEIN"

/protein_id="CAD15032.1"

/db_xref="GI:17428345"

/translation="MSANDAPAVKPSLLTLEAALQLLAAARGLTETETVPTLHANGR

VLAQAVRSLGNPPADNTQMDGYAVRAADITAPGTRLKVQRIQAGQVGRPLQAGAA

RIETGALIPGDADAVVMQEQCRPDPTGTVDHVPAGEMTIRAGEDIYTAGAEILPR

GTRLCAQQLGLAASVCAGLOVVRPRVAVFTGDELAAMPGLPKGALYNSNRFTLR

ALLENLGCPTDFGIVPDTLQATROTLEAAEAHDLITSGSVSVEEDOHIKPVEAE

GRNLNQIAIKPKGPLAFGEVRRPQGMAGGPAFFIGLPGNPVPSFTVLLPVRPIL

RLUGVDDVTPKRIIPRADPALKDORRNEFLRINAGGGLDLFPNQSSGLTSTVMG

GDLDNPNPQPIAAGDTVAFLPFAVL"

6310..6573

/gene="moad"

/note="RSC1331; RS02853"

6310..6573

/gene="moad"

/function="small molecule metabolism; biosynthesis of

```

Qy 378 eserGlyGlyLeuPylValArgAspAlaGlyLysLeuAspArgGlyLysPheArgGlyLysLeuAlaValPheHisMetGlyAspAla 379
      ||||| ||| ||||| |||||
Db 102308 GTTCTTCGGCATCGGTAC-----GAGCGCAGGACCATCGACTGACCTTCCAAACAGCGC 102361
Qy 398 -----GluPheLeuAlaGluGlyArgGlyLysIleProGlySerAspIleAspLe 413
      ||||| ||||| ||||| |||||
Db 102362 GTTGGCTACCAAGNACTACGTGGCCAGAAAGCGCGCGNT----- 102401
Qy 413 uGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnVa 433
      ||||| ||||| ||||| |||||
Db 102402 -----TCGACAACTTCGGATCACCATCGGTGGTCGAAGGACCGCGGCACAGCGC 102454
Qy 433 lLeuHisProGluAsnGlyHisTrpLeuAspGlyLysIle-----GlyTh 448
      ||||| ||||| ||||| |||||
Db 102455 CCTGGTGCCAGCGCGCGCTACCAAGCAGGCCAACCTGGAGTTCGGCATTCCTCGGCGG 102514
Qy 448 rThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTy 468
      ||||| ||||| ||||| |||||
Db 102515 CGACCTGCAGTACTAC-----CGTGGCTACTA 102541
Qy 468 r-----PhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyG1 485
      ||||| ||||| ||||| |||||
Db 102542 CCAGCACCACTACTCTACCGGTGTCGAAGTCGTTTC-----ACGCTGGCGTTCAACACGA 102598
Qy 485 nAlaGlyTyThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 505
      ||||| ||||| ||||| |||||
Db 102599 GATCGGTACGGCCACCGCTACGGCACACAGGACTTCCCGGTCTTCAGAACTACTACGC 102658
Qy 505 rGlyGlyAlaSerSerValArgGlyTyTrpGluLeuAspSerIleGlyLeuAlaGlyProAs 525
      ||||| ||||| ||||| |||||
Db 102659 CGGCGGTATCGGCTCGGTGCGCGGCTACGAGACGACACGCTGGGCCCGCGCACGCCAA 102718
Qy 525 nGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyTrpGlnLeuProPh 545
      ||||| ||||| ||||| |||||
Db 102719 CGGCGTGCCCATCGTGGCGGCACCAAGTTCTGTCGGCAACATGGAATTCATCTCTCCGCT 102778
Qy 545 e-----ThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAl 560
      ||||| ||||| ||||| |||||
Db 102779 GCCGGGCTCGGGCGTGAGACGACCGTCCGCTGTTCCACCTCTCTCGACTACGCGCAAGT 102838

```


Db	200988	ATCCAGGTATCGAAGCAACGACCGCTATTTCGGTTTCGGCGGACAGGTCTCGACCAACG	201047	
Qy	311	TyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIle	330	
Db	201048	GACGGCTTCGGCTCCAGGCTATTGGGCGCATCTCTTCGGACGTCGCGAATCG	201107	
Qy	331	GlySerValValTrpAspMetAspLysTyr---GluThrThrLeuAlaAlaGlyIleSer	349	
Db	201108	CTTCGGATCGAAGGCTCGGTTCGATCGCTCGGCGAGACCAACCGATGTCGCGCGCTCGAT	201167	
Qy	350	GluProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThr	369	
Db	201167	-----	201167	
Qy	370	ThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAla	389	
Db	201168	-----TATTCGGCGGCAATCTCTTCGCGAACCAGCGGCC	201203	
Qy	389	-----	389	
Db	201204	TTCCGACCGGCTCGACCTTTACCGCAAGCGTCAAGCGCGGATCGTCGATCCGGATGCC	201263	
Qy	390	-----GlyIleAspAlaArgLeuGlyAlaGluPhe-----	399	
Db	201264	TATAGTCCCAAAACGGTTACCGCGCGCGCGCGCCCTTCGAGCTTTCGCCCGCAAGAC	201323	
Qy	400	---LeuAlaGluGlyArgLysIleProGlySerAspIleAsp-----LeuGlyAsn	415	
Db	201324	ACGTTCTCCGTCGGCGGGAAGTGGGTGGCGGATGTCGACGACGCTTTCGGCTCGAAT	201383	
Qy	416	SerHisAlaThr---MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu	434	
Db	201384	TCCTATATCACCGCGCGCCCTCCCGTTCGAATATGTCGCGCACCGCATGACAAGCTG	201443	
Qy	435	HisProGluAsnGly-----HisTyrLeuAspGly	444	
Db	201444	AACCCGAGGAGGCTACAGGGCGCTGATCAATGCCAAGCGAGCTACGAGATCGAGGGA	201503	
Qy	445	LysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla	464	
Db	201504	AAG-----ACCTTTTTCAGCTCC-----TTCGAAGCTTCCGCG	201536	
Qy	465	ArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly	484	
Db	201537	TCCGGCTATTACGGCTTTGGGACCGAAAAGCG-----TTCGTGCTGCGCGCG	201584	
Qy	485	Gln-----AlaGlyTyrThrValAlaArgAspAsn---AlaAspValProSerGlyLeu	501	
Db	201585	AAGTCGGCGCGCGGTGCTCGTGGCGGACGAACCTCCGATATTCACGCGCACCCGG	201644	
Qy	502	MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu	521	
Db	201645	CGGTTCTTCCTCCGCGCGCGCGGTCTCGTGGCGGCTATTCTTCCACGAGGATCAGCCCG	201704	
Qy	522	AlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGluTyr	541	
Db	201705	CGGCACCGCATGACGAACCTGACCGCGCGCGCTCTCTATGTGACGGCTCCCTCGAAGCC	201764	
Qy	542	GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisaspMetGly-----	558	
Db	201765	CGATCGCGCTCACCGATAGATGCGGCGTCTGCGCCCTTCATCATGCGCGGTTCG	201824	
Qy	559	---AspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyVal	577	
Db	201825	GATAGCACAGCGCGCATTTCTCC-----GACATCCGGCGCGCGCGGCTCGGTCTG	201878	
Qy	578	ArgTrpPheSerProLeuAlaProPheSerPheAspIleAla	591	
Db	201879	CGTATCGCAGCGCATTCGGGCGCATCCGCTCGACTTCGCC	201920	
RESULT 24				
AE006218/c				
LOCUS	AE006218	10302 bp	DNA	linear BCT 08-MAR-2001

DEFINITION	Pasteurella multocida PM70 section 185 of 204 of the complete genome.
ACCESSION	AE006218
VERSION	AE006218.1
KEYWORDS	GI:12722230
SOURCE	Pasteurella multocida.
ORGANISM	Pasteurella multocida
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Pasteurella.
AUTHORS	1 (bases 1 to 10302) May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and Kapur,V.
TITLE	Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE	21145866
PUBMED	11248100
REFERENCE	2 (bases 1 to 10302) Zhang,Q. and Kapur,V.
AUTHORS	Direct Submission
TITLE	Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
JOURNAL	University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES	Location/Qualifiers
source	1..10302 /organism="Pasteurella multocida" /strain="PM70" /db_xref="taxon:747" complement(70..1830) /gene="PM1809" complement(70..1830) /gene="PM1809" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03893.1" /db_xref="GI:1272231" /translation="MKLSNTKASLTITKTVFFIGSVLLSTVSLAQQLELVQKGIQNTQ LNNVRIYVGMIDKEADGSRHKQVREAIKALRAYGYVQSEVFQIESQKPKKD OLIAHVIGKPKVIAADVLOGEATQDPETAVTKNIPEKGSILNHETDYNKSHIQ KIALSGYDGDPLVSRLEVRPSTQAAWRLDYDGERVRFGEVTFENAOIREDYLRN MINFKQGYPLINDLSLTNNVTSNWFSSVLMQPVLDDEHKIYVNDVLLQPKKNSM EVGIGWASDVGRPLQGWTKPWINNHSFRNLVVSAPKQTLTEATYKMPLLKNLNY LYFYSAGLENENKNDTESPASSLSAIRYWNHEAGWQHSGLGRVRSDFIOANVKDTL HRTVTSRTRKQGLFTTGTQKLTIDLGRWMLSDVDFLKMGGSLKMGSLVRYTQLH TGSFEYGVYVPMWLATPADTCLAAQFTTRELRYGAGMGVRAWSPVGAIKFDIATP VRDKNSKNIQFYGLGAEI"
gene	complement(1980..2609) /gene="narp" /note="PM1810" complement(1980..2609) /gene="narp" /codon_start=1 /transl_table=11 /product="Narp" /protein_id="AAK03894.1" /db_xref="GI:1272232" /translation="MSEKMKVLLIDDDPLMRGKIKOLMELAPNFVAVADAGNCGSIS LALQTPDLIIIDLNKMGSLGSLDTLALRAEGIDARILILTYSDLSKSDIFTLLDAGAD GYLKQTEPDALLSQIRIAQGEVILSDSIKDALIEROMNADPIYSITDREMVDVLRLI ATGLSNKQIAGQLFISEETVKVHIRNLRLLRKNVHSRVAATVLYFEHGR"
gene	complement(2625..5141) /gene="cyaA" /note="PM1811" complement(2625..5141) /gene="cyaA" /codon_start=1 /transl_table=11 /product="CyaA" /protein_id="AAK03895.1" /db_xref="GI:1272233" /translation="MNYDLFSAOKKVEYLDKLRILRALSGSSGSEFOHVQOLLTLHLHI NHPNLPGYVADAPVGIADPVISPYQKQYLLTTVTVPSEANQSLPSPFSYRSTNAILGY
CDS	


```
Qy 254 LeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 1143 ATTAATGATCTTTTACCTTAACCAACAATTTACACGTCGTCCTCAAAATTTGCTTGTAGCTCTGTT 1084
Qy 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
Db 1083 TTAATGCAACCTGTTTGTAGATGAA---GAACACAAAATTTGAATGTCGATGTATGCTC 1027
Qy 294 ThrGluValLysArgHisLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
Db 1026 CAACCACGCAAAAAATAGCATGGAAGTGGGATTCGCTGGCATCTGATGTGGACCT 967
Qy 314 GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerVal 333
Db 966 CGCTTGCAATTAGTGTGCACAAAGCTTGGATTAACAATCGTGTGCACAGTTCCGTACC 907
Qy 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGly-----Ile 348
Db 906 AATTATATGTTTCTGCGCCAAAACAGACCTTAGAAGCCACTACAAAATGCCCTTTATTG 847
Qy 349 SerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 846 AAAAACCGCTCACTATTATACGAGTACTCGCGGGTTTAGAAAATGAAAATAAAAAAT 787
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArg 388
Db 786 CACACAGAAATCCTTTCGCTCTCCTTA-----TCTGCAATTCGCTATTGCAATCATGAA 733
Qy 389 AlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 732 CGAGGTGGCAACATCTTTAGGCTTAGTGTCCGCTATGAC----- 691
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 690 TCTTTTATCAAGCGAATGTAAGACACAAACATGCTGTGTCCCACTGCCTCTGTC 631
Qy 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 630 AGACGAACCCGTTTACAAAGCGGGCTATTCCCGACTTGGGGGAT-----ACACAAAAA 577
Qy 446 IleGlyThrThrLeuGlyThr-----PheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 576 CTCACCATCGATTTAGGTGCTACATGGTGGCTGCTGATGTTGGATTTCCCTCAAAATGCAA 517
Qy 464 AlaArgAlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThrPheIleIle 482
Db 516 GGGTCGAGTTTATGGTCAGAACTTATTACAGCACCATCTGATC-----GTTACC 466
Qy 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
Db 465 CGACTAGAACCTCGGCTGGTTACATACCAAAAATATCGAACGCATTCACCCGCTTGGCT 406
Qy 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 405 TTTTGTGGGGGGAGATCGCAGTATTCGTGTTATGCTGTATGCTGTATAAAAAATTCGCCCAAAA 346
Qy 523 GlyProAsnGlySerValProGluArgAlaLeuLeuValGlySerLeuGlyTyrGln 542
Db 345 AACAAATCGAGGCAATTTAGTTGGTGGCTCTGCTTATTAAACCGGTCATTTGAATATCAG 286
Qy 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAla 562
Db 285 TATCAAGTGTATCTGACTGTTGTTAGCCACTTTTCGCGGATACGGGACTGCTGCTAAC 226
Qy 563 AsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
Db 225 CAGTTTACAAACAAAGAAATACGCTATGCGCCAGGAATGGGGGTACGTTGGCGCTCACCC 166
Qy 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAsp-----LysLys 598
Db 165 GTAGGGGCAATTAAGTTTGATATTGCCACCCCGCATCAGATAGAAACAGTAAAAAT 106
Qy 599 IleArgTrpHisIleSerLeuGly 606
```

```
|||||.....| | | | |
Db 105 ATCCAATTTTATATTGGACTAGC 82
RESULT 25
LOCUS
DEFINITION
Agrobacterium tumefaciens str. C58 circular chromosome, section 233
of 254 of the complete sequence.
ACCESSION
AE008175 AE007869
VERSION
AE008175.1 GI:15157816
KEYWORDS
Agrobacterium tumefaciens str. C58 (Cereon).
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE
1 (bases 1 to 10029)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 10029)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
Location/Qualifiers
source
1..10029
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/strain="C58"
/db_xref="taxon:181661"
119..2044
/gene="AGR_C_4742"
119..2044
/gene="AGR_C_4742"
/notes="conserved hypothetical protein PA2543 {imported}"
pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_C_4742p"
/protein_id="AAK88337.1"
/db_xref="GI:15157817"
/transl_table="MRTNPKTGIAYRKAGTALAVALTFALYPAPARDAFAKPLFG
MRUWGSEPEVEVINPVKYAVTDAADADKSLKSSLESSLADKDKPASGDLGLLI
KARRDRRLIAALYENARYGIVNVTAGKNVDLPNVPVHDSTPVPVITVTPGPK
FTLGNVRLEGDVTGRNLDVGLIAGGDAGSLAIIAGNKLIDDLKAEGRPLAKLTRE
AVANHATNTVDITMAAEGCPVAPLGAVTGTGVDGDFIRRYSLRNGGSPYSPEKLR
KAADRRLQGVFSLSLTIKEAGTLIAROGTITPLEVSEGRHRYFGVGAQYSTTGIGLQ
GYGHRNLFQGAESLRIEGSVAIEASSVEGMDYSAGITFTKPGMFPNPTTKSLI
AKTENPDITYRAKTLITGTAGFAYELNDTDAAGLEIYOWADTEDAFKNEYLTTSILE
FVRDTRDDKLNPTGFRASLAAPSYEALNFTFSFEGSITGYKGLGADRILIMAGK
LSGVLVGGSDLODIPTRRFAGGGSVRGYSYQETSPYNAAGDATGGRSYVVGVE
ARIKVVDITGLVPFDAGVVDGVTDFDIRAGAGIGLRYATPFPGLRLDVAEMPLEK
YDKGNFPIYAGIGQSF"
2087..6250
/gene="AGR_C_4743"
2087..6250
/gene="AGR_C_4743"
/notes="(AY010120) putative periplasmic protein"
/transl_table=11
/codon_start=1
/product="AGR_C_4743p"
/protein_id="AAK88338.1"
/db_xref="GI:15157818"
/transl_table="MMKTLIRLLKWLGYAALCCVLLVLLAVLTVFGVTPMCAARTAAAOI
SLVSTPDQTIETISPPGGLLTLGRLDNVTLSDRQGYARLNQIADVMSPLSLIAGTF
HADHVSAGSIDVERQPLPAQQTTSKSGGSLPIEIVDNFSPDLSLGGSLGRFLA
LTAEGNLKAQDDMRSLSTAHRNRNAADVAVFLPNENVLKLAEMKEPGLGLIA
TLISLPTPAVAIDLNGEGPLSSWTGLRCNVAGNAVNVNTRGHLDDCQDTRRIETAG
GGOPDLLPPAFRLQGAETKLDANATLYPQGRIEIGSSSTLETGALLTTASGVDPNG
QNDLAANLGTAGCPVDFRWPPLNGEIQALINGLDLSLKGGAADAHNLNTTVSLRSLALP
QGRDIDDKLTAESADLNTNRSGTITATVLSVASSFSVDILRPLTIKAPLSLT
SSAIGFNGATLESASVGGTGLNFDLADNRNLTSSVQLFALPATLPPALAEKFDFTTAL
```

QGDIDLTIGGRTSIENLYIKSGTIEAAGNVSLNDALSADLKGZFPALAKLTPPOAKGI
ADFAIEASALAAFPVNTVTSSESAIAGRLKALKVNASGKADPAAPAKFTAGSGL
DROKIDANAVQVETNGTAIPELHVAAGNINLKGLOFSQQFLPTNGSNFNPDLALL
ALAAQADGDIAGDIALSNANGKIAATIKANGNIRQGTITISKLADISIDDIKAL
ALNGKISADVWAGTAALISGLNATIGHSTTTFADNGRIDNAPLVKGSADTGSPM
TVYDITFAAPKIPVIRLEKPSITIAQNGTARIADLTITIGDGRVEVNGTAGSTLDIK
ADIRSLPASLANAFAAGLDAAGISGTVSAKGAADSDPSVDYNLNWANAFAQTRAAGL
AALGKANGRFAGGNLQIDTDTVGCGNSLGGSGISGAGNRPISMAFSGRLPFSAVA
ACTAAGLVDVCTAIDVNIAGTASAPVGTSTITTDGRLTDVNRNLITINGLATIKF
DRDAVIRSLTKLVGGGTISCTGSGVGTAGSGFPADISITLDKAGYNDGTLVTVVS
GLTLKGPLNSPALGNGNLTDRSAITIEPKLPASLTEDIVKHNAPKPRVQAQALG
ADQGGSGSTINLDQVNWAPSGVIFVRGIDDAELTNGLTITIGTAAPVVISGFFEMR
GQELTLTTRDTNGITFGGLVPLDMKADSTVSTTVTVSVSGNANDPTFAFSSA
PALPQDEYMAQILFQSGMSKLSALQIARLADAQAOLAGRSTSLFDKLSNGLVDLDO
ISTDSEGOARVLSAGKYLNBRTYLEQSGESCAKAIINLDVGRGVKLRCEAGNGEGA
AGIFVEKY

gene

6302..6895

/gene="AGR_C.4744"

CDS

6302..6895

/gene="AGR_C.4744"

/note="(AF312878) chemotaxis protein CheW2"

/codon_start=1

/transl_table=11

/product="AGR_C.4744p"

/protein_id="AAK88339.1"

/db_xref="GI:15157819"

/translation="MKTEMVFPVAVQAGRTPNHMQVSVHLRPPORALGRILMAWIN
SNFGGETLITAFRLHDOFCVKTTTIREIRGWAPSTPIPHAPKDVIGVNLRSYI
PIIDIAHKGMSVIANERSALIVAEVHNKVIIGMLVDVRVSDILTIPANOVQVPETISA
SFDKSYSGEIVRANSEMICFLNLAKMFKGTEADLAA"

7187..9004

/gene="AGR_C.4747"

gene

7187..9004

/gene="AGR_C.4747"

CDS

/note="putative mcpA"

/codon_start=1

/transl_table=11

/product="AGR_C.4747p"

/protein_id="AAK88340.1"

/db_xref="GI:15157820"

/translation="MGAASRPWTIDWITMLGLKGSADNRNMLDIAISRQAVFEDLK
GNILNRFNCMALGYDLAEIVGKHHRIFCDREIAASHAYQEFWESLARGEFOAKERY
RIRKDSAVYIEASNPVPSGPKYKVVKIATDITAKRIKAVEDAGKLEALSRSQATI
EFPDGTITANPNFCATVYDLKEIEGKHHRMFCDPAYTASAAVNFWRLASGEFI
SDEFVYKNGKEINIOAAYNPVLEAGKVVVKVATDVTPRMSAITVLADALQALA
AGDLVRLDTPSPVSMELRADFNVEVVGKLOATMOTISHNASTIASGSGEIRIADQL
SRTQQAASLETAALAEITITVDASORAAEAGKVLIRTFREAHSGDVVQQAIS
AMDASRSGEITNIGVIDDIAFOTNLNLAAGVEAARAGEAGKGFAYVAQVEVELA
QNSAVAKEIKSLINTSREQVANGVDLVGRGTGALRDIQSOGEIDINVSALVEASRE
RANGLKEINOAVNMVDQATQKNAAMVEETTAASHGLASEADNLYLLRQFKISGDQOI
SSAOPITDKKAASISRLPTQPRVATARAHNAAEWAEP"

complement(9035..9880)

gene

/gene="AGR_C.4748"

CDS

complement(9035..9880)

/gene="AGR_C.4748"

/note="(AF298190) NAD-dependent formate dehydrogenase

subunit C"

/codon_start=1

/transl_table=11

/product="AGR_C.4748p"

/protein_id="AAK88341.1"

/db_xref="GI:15157821"

/translation="MWCVFRUGMIVSVSEAYSRVARTALRECTPATGERAPVEEV
PVAFSGGTHAVMMASPADLIDFVAGFSLTEGIIISMRSEIQAEIVVEAGQDIDVQVD
LTDADALARRRRHAGPVGGGIESIEQAVRVVPDLSSVSKSVHNDIAQVAVKA
LNDADTLNRETRVAGGFPNPREGLLAVREDVGRHNDLKLAVAGVNIAGMIGV
AVTSRLSVENVOKTALGCPVLAAISAPTALAIETADRAGITLVALVRGDFEITRA
DRILL"

BASE COUNT

2195 a 3131 c 2814 g 1889 t

ORIGIN

Alignment Scores:

Pred. No.:

5.89e-12

Length:

10029

Score: 323.50 Matches: 148
Percent Similarity: 36.95% Conservative: 94
Best Local Similarity: 22.60% Mismatches: 290
Query Match: 10.20% Indels: 123
DB: 1 Gaps: 21

US-09-857-669-2 (1-609) x AE008175 (1-10029)

QY 6 ThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyrAlaProAlaAlaAsp 25
Db 173 ACCGCGCTTCGGCTTGGTTGACGCTTCGCCCTGTATCCGCGCTTT-----GCCCGTGAT 226
QY 26 LeuSerGluAsnLysAlaLaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 227 GCCTTCGCTTCAACATGTTTCGCATCGCGCTCTGGGTTTCGGAAGAACCCGGAGTGCAG 286
QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGlu----- 63
Db 287 GTTATCAAC---CCGGTCAAAATATGCGTTCACGCTTACGCGCGCCGACGCGACAGTCC 343
QY 64 IleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeu 83
Db 344 CTGAAGACGAGTCTGGAACACACGCTCGCTTCTTCTGCGCCACAG----- 388
QY 84 AspLysGluGlnThrGlyPheLeu-----AlaGluGluAlaProAspAsn 98
Db 389 GACAAACCTCGCTCCGCGGATCTCGGCTTCTGATAAGCGCGACGACGACGACAGG 448
QY 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 449 CTGATCGCGCACATTTATGAAACGCGCTTACGCGGCGCATCGTCAACGTCGACCGTGCC 508
QY 117 -----GluLysAspGlyAla 121
Db 509 GGAAGAACCTTGATGACCTGCCAACCCGCTTTCGATCAGACGACGCGCTGTACCC 568
QY 122 TyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIle 141
Db 569 ATGCTCATACCGTAAACCGCGCGCGCGAATTCACCTTCGCGCAAC-----GTTGCGCTT 622
QY 142 LeuGlyAspIleLeuSerAspGlyLysLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161
Db 623 GAAGGCGATGTA---ACGGCGCCCAATCTTGATGAATACGCGCTGATCGCC----- 670
QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
Db 671 -----GGCGGTGATGCGCGGTCTCTGCGCAATATCCGCGCGCGCAACAG 715
QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyLysThrArgAla 201
Db 716 CTGATCGATGATCGAAGCGGGAAGCGCGCGCTCCGGAACACTGACGAAACCGACGCGG 775
QY 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 776 GTCCCAAAACACCGACGAGTAATGCGTGCACATTACATGCGAGTGAAGGTGGCCCGCTC 835
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 836 GCGCGCGTCCGCGCAGTCACCGTCACCGCGGCAAAACCCGTCGATGCGCGCTTCATTCCG 895
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGln 261
Db 896 CGTTATTTCGCGCTTAATGCGCGGAGCGCTTATTCGCGGGAAGAACTGCGCAAGCGCGCC 955
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 956 GACCGCGCTCGCCAGCTCGCGCTCTTTCCAGCTGACGATCAAGGAAGCGGGAACGCTT 1015
QY 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1016 GCGCGCGACGCGACCATTCGCTCACCATCGAAGTCCGGAAGGACGACCGCTATTATTT 1075
QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321

```
Db 1076 GGCCTCGCGCGCAATATTCACGACCAAGGATATCGGGCTTCAAGGATATTCGGGTCAC 1135
Qy 322 TyrAsnLeuPheAsnLysGly-----TyrIleGlySerVal----- 333
Db 1136 CGTAACTCTTCGACAGCGGAATCACTCGGCATCAAGGCTCTGTCTCCGTTATTCGG 1195
Qy 334 -----ValTrpAspMetAsp----- 338
Db 1196 GAGCATCGAGCGTCGAAGCATGGATTATTTCGGCGGGCATTAACCTTCACCAAGCCGGG 1255
Qy 339 -----LysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg 352
Db 1256 ATGTTCAACCCCGCACGAGCTTCAAGCAGCGCTGATGCCAAGACCGAAACCCGAC 1315
Qy 353 AsnTyrArgGlyAsnTyrTTPTrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
Db 1316 ACCTACAGGCGCAACCGTACGCGGACG----- 1345
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIleTyrPyrValArgAspArgAlaGlyIleAsp 392
Db 1346 -----GCGGGTTTCCTCCACGAACCTGAACGAT----- 1372
Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
Db 1373 -----ACGGATACAGCGCGCGGGTGTGGAGTCCAAATGGCGCATACCGAA 1420
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
Db 1421 GATCATTCGCGCAAAACGAATATCTCACGACCTCGATACCGTGGAAATTTGCGCGCAC 1480
Qy 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThr 452
Db 1481 ACG-----CGGCACGATAAGCTTAATCCGACCGAAGGTTTT 1516
Qy 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro 472
Db 1517 CGCGCATCTTGGCGCGCAAGCAAGCTATGAAGCGCTCAATGGCAGCTTCTTCTCTCT 1576
Qy 473 GluAsn-----LysLysLeuGlyThr-----PheIleIleArg 483
Db 1577 TTCAAGGCTCTATCACCGCTATTAAGGCTTGGTCCGCAAGACCGGTTGATCATCGCA 1636
Qy 484 GlyGln-----AlaGlyTyrThrVal---AlaArgAspAsnAlaAspValProSerGly 500
Db 1637 GGAAGCTCTCGCGCGGTGTCTGTGGCGGCGCATCTACAGGATATACCGACCCACC 1696
Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 1697 CGCGCGTTTTTCGCGGTGGCGGATCGTGGCGGCTATAGCTACCGAGGATTCG 1756
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 1757 CCCTATTAATCGCGGCGGATGCGACTGCGCGCGCTCTACGTGGTGGCTCGCGTTGAG 1816
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGly----- 558
Db 1817 GCCCGCATCAAGGTCAACCGATACCATCGGCTTGTGCGGTTTTTCGACGCGCGCTCGT 1876
Qy 559 -----AspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGly 576
Db 1877 TCTGATCGGTTAACACCCGATTTCTCC-----GATATCGCGGCGGCTCGCGTATCGGC 1930
Qy 577 ValArgTrpPheProLeuAlaProPheSerPheAspIleAla 591
Db 1931 CTGGTTACGCGAGCGCTTCGCGGCTCTCGGCTCGATGCGCC 1975

RESULT 26
LOCUS AE009208
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section
234 of 256 of the complete sequence.
ACCESSION AE009208 AE008688
```

```
AE009208.1 GI:17741110
Agrobacterium tumefaciens str. C58 (U. Washington).
Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 10301)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Zhou,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
21608550
11743193
2 (bases 1 to 10301)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA
Location/Qualifiers
1. 10301
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
83..1276
/genes="dxr"
/notes="Atu2612"
83..1276
/genes="dxr"
/notes="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"
/protein_id="AAU43593.1"
/db_xref="GI:17741111"
/translation="MTNASEMPKRLITLSTGSGTNTLDVVRQLGRRDGFETMLTGTG
AGNIALLAQAEFGAQLAVTDDDKYEALKSALACTGKVAAGMAGLEAASMDAGW
VMAA IAGTPLGAPLTAAARGADIALANKECLVSAGDVFRTVKQGGRLIPVDSBHS
ATFOCLTGEYKQKNAVERIVLTASGCPRTWSDMSNVYADIAAHPNWSGLKVSIGS
ASMFNKGLEMIENAKYLDLPQDVIVIPQSIHSHWGTIDGYSYIAQJGSPDMRTAI
SVALTYPERNLSVERLDFAKLALDFEAPDEARFPALRLARMLERGGJGQAALNAA
ETAFHAFVAGGIGFLDMAEIVETVMDRHDGRTAETMDVFSADSEARRHLELIAT
KEKAA"
complement(1312..2355)
/genes="hema"
/notes="Atu2613"
complement(1312..2355)
/genes="hema"
/notes="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="5-aminolevulinic synthase"
```


QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAla 201
Db 3924 CTGATCGATGATCGAAGCGGAGCGCGCGCTCGCGAACTGCGAAACGCGAGCG 3983
QY 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
Db 3984 GTCGCAAAACCGACGCGAATACGGTCGACATACAAATGGCAGCTGAAGGTGGCCCGTC 4043
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 4044 GCCCGCTCGCGCAGTCACCGTCACCGGCGGAAAAACCGCTCCATGGCGACTTCATTCGC 4103
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 4104 CGTTATTGGCGCTTAATGGCGGAGCCCTATTTCGCGGAAAAAGCTGCGCAAGCGGCC 4163
QY 262 GlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 4164 GACCGCTTCGCGCAGCTCGCGCTTTTCCAGCGCTCACGATCAAGGAAGCGGAAACGCTT 4223
QY 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 4224 GCCCGGACCGCACCATTCGCTCACCATCGAAGTGTCCGAAGCAAGCACCGTTATTTT 4283
QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321
Db 4284 GCGCTCGCGCGCAATATTCACGACGCAAGGTATCGGGCTTCAAGGATATTTGGGTTCAC 4343
QY 322 TyrAsnLeuPheAsnLysGly-----TyrIleGlySerVal----- 333
Db 4344 CGTAACCTCTTCGCGACGCGGGAATCACCTCGCATCGAAGGCTCTCTCGCGTATTGGC 4403
QY 334 -----ValTrpAspMetAsp----- 338
Db 4404 GAGCATCGAGCTCGNAGCATGATTTATTCGGGGCGATTACCTTCACCAAGCGGGG 4463
QY 339 -----LysTyrGluThrThrLeuAlaGlyIleSerGlnProArg 352
Db 4464 ATGTTCAACCCCGCAGCAGCTTCAAGACGAGCTGATCCCAAGACCGGAAACCGGAC 4523
QY 353 AsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
Db 4524 ACCTACAGGCGCCAAACGCTGACGGGCACG----- 4553
QY 373 LeuGluLysArgAlaPheSerGlyGlyIleTyrTrpTyrValArgAspArgAlaGlyIleAsp 392
Db 4554 -----CGGGTTTCGCTACGACTGAACGAT----- 4580
QY 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
Db 4581 -----ACGGATACAGCGCGCGGGTCTGGAGTCCCAATGGCGCGATACCGAA 4628
QY 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
Db 4629 GATGCAATCGGCAAAACGAATATCTGACGACCTCGATACCGCTGGAATTTGCGCGCAC 4688
QY 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThr 452
Db 4689 ACG-----CGGACGAGTAAAGCTTAAATCCGACCGGAAGGTTT 4724
QY 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPheThrPro 472
Db 4725 CGCGCATCCTTGGCGGCAAGCAAGCTAATGAAGCGCTCAATGGCAGCTTCTTCTCTCT 4784
QY 473 GluAsn-----LysLysLeuGlyThr-----PheIleIleArg 483
Db 4785 TTCAAAGGCTCTATACCGCGCTAATAGGGAAGTGGTCCGCAAGACCGGTTGATCATGGCA 4844
QY 484 GlyGln-----AlaGlyTyrThrVal---AlaArgAsnAlaAspValProSerGly 500
Db 4845 GGAAGCTCTCCCGCGGTCTCTGGTGGCGCGGACGATCTACAGGATATACGACCAACC 4904
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520

Db 4905 CGCGCGTTTTTCCCGCGTGGCGCGGATCGGTGCGGGGTATAGCTACCAGGAGATTCG 4964
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 4965 CCCTATAATGCGGCGCGGATGCGACTGGCGCGCTCTACGTGGTGGCTCCGTTGAG 5024
QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGly----- 558
Db 5025 GCCCGCATCAAGTCACCGATACCATCGGCTTGTGCCGTTTTTTCGACGCGCGCGT 5084
QY 559 -----AspAlaAlaAlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGly 576
Db 5085 TCTGATCGGGTAACACCGCATTTCTCC-----GATATCGGGCGCGTGGCGTATCGGC 5138
QY 577 ValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla 591
Db 5139 CTGCGTTACGCGACGCGCTTCGCGGCTCTCGCTCGATGTCGCC 5183

RESULT 27

AE003957

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AE003957 10074 bp DNA linear BCT 15-JUN-2001
Xylella fastidiosa 9a5c, section 103 of 229 of the complete genome.
AE003957 AE003849
AE003957.1 GI:9106207

Xylella fastidiosa 9a5c.
Xylella fastidiosa 9a5c.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
1 (bases 1 to 10074)

Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hohensei,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.,
and Marinho,C.L.

The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
Nature 406 (6792), 151-157 (2000)

20365717
10910347
2 (bases 1 to 10074)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hohensei,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,P.,
Lemais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Laigret,P.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marinho,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr.,A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira,
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa,Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva,
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,

RWYDRAKQFALHGLHPWRNNGKLTLOAQELEIGTLLDSVTRHSGAVENLNLQADV
RHAMALTLLHGYYQDWRMRGDSQLRIETAPKQANARPAQVFNQVFTLSATC
LDAOGGALCLHMDHPGELVKLGKQPLTILVHPWLPKNGRSNLRGALTLDGHLRP
HSDAWESSFRASSLEGIIRLGEPRYADVLGVQDQDELVRVDFQSVQDVLQSHITVCKL
GIGFQNGFVDKAFWQGWDPYAPILRTGDLVMNNSRLYLEMFPVQDGNPRLGIECHVSL
RGTSPGLQDRLDFTFAFVYPMGLNLQSQGRFTAPQDGSARLILASVRSVGGTLL
IDGGLSGFENNPLQNTIGRNLVAYLTSELRVYSPDLORFLGRLNTMYLOGEVAVSE
ADTLDERLQSGVSDVQVQDVPDPTAYTSPDMLNRLVLGDKVMNSGVLKGGLS
GOMQIRARPCHEWTAANGDLVRGRYKAYGQDILTIRGOLVMNNIVSDPVRNTRAOPR
IADITAGIDVSGRAQSPVQVQDVMVQSGSEALSYLIRGKRLGTASDEBQVQVSAANA
ASAGGELTALQALGAKLQDLDAGISQSTLGGSVLGFYGLFKYLPKLYGVYGVQVMSVGS
VLTKYLLSRGFLVESSTVTKGSVNNRKE*
complement (6826, 6945).

complement (6826...6945)

ADIDJELKURGVYSGEVLVDUPERKATYVLEDMNNIYUGKRVMSGGLRGKGV
GOMQIWRARCHEWNTAGVLDGVRGYKAYGODTLIRGLQVMSNNIVSDPRNTRADRR
IADIITAGIDVSGRAQSPRVQWSDPWNQSSEALSYLVLRGLKGTASDEWQVVASAAA
ALSAGGELLTASOLGAKGLDLDAGISQSTLGGSVLGFCKYLSPLKYGVYMWVSGSGS
VLTKYLLSRGFDEVESSVETKGSVNRREK"

VLTLLSRGFDLEVESSTVETKGSVNWRKE*

complement(6826...6945)

```
complement(8826, .0945)
/gene="XF1233"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
```

```

/transf_cable-11
/product="hypothetical protein"
/protein id="AAF84043.1"

```

```

/protein_20= MG040433.1
/db_xref="GI:9106211"
/translation="MFLFEHEHIGNSAMSLAIGASFLLVDQPGFIGQAMRCKP"
7500..8396
/gene="XF1234"
7500..8396

```

```

/500. : 8330
/gene="XF1234"
/note="similar to SP|P77541 (percent identity: 62 %/query

```

alignment coverage: 98.3 %/subject alignment coverage: 99.0 %): identified by sequence similarity: putative: ORF

```
located using BIMMER/KSfinder
/codon_start=1
/transl_table=11
/product="carboxyphosphonoenolpyruvate phosphonmutase"
/protein_id="AAF84044.1"
```

```

/db_xref="GI:9106212"
/translation="MSTSLPIQTPGQCFFALTMTERPLQVIGAMNAYHALLAQRAQYR

```

ATYLSGGVGAAGSGLPDLGMNTLEDVLIDVRITTDVCLPLLVDTDTGEGSPALNTN
 RTKTLIKAGAACACHIEDQVAAKRCGRHPEKIVSLSDMARVKAADARTDANAFLLI
 ARTDIAQEGVSEAIERAGIVAEAGADSIIFAEAVSDLETIRHFVERIRVPLVANNTEFF
 GKPTLPTLRHAFEGVGAIVLPLFPLSAFRAANKAAETVYNAIRDDGQQQVLPVHMOTRAE
 LYERLDYHAEQRLDALSARST
 complement (8703 . 8778)
 /gene="trna-Lys-1"
 complement (8703 . 8778)
 /gene="trna-Lys-1"
 /product="trna-Lys"
 /note="found by TRNscan"
 9291 . 9446
 /gene="XF1235"
 9291 . 9446

9291. : 9440
/gene=XF1235"

/note="hypothetical protein; identified by sequence similarity: putative: ORF located using Glimmer/RBSfinder"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84045.1"

```

```
/db_xref=GI:9106213#  
/translation="MLPLAGSGTTGVAALRAGHRFIGMELSLGYSDVAKQRLPDRVCG
```

```

FIRDVYL"
9570..9887
/gene="XF1236"
9570..9887
/gene="XF1236"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA894046.1"
/db_xref="GI:9106214"

```

```
/translation="MIMIRKTLRRIAAYLAYLPNMRLLEVVMITALVALAAHVLWS
VNHEKLTORATHIOEOVASMERTLNEKAAMNAEODAAIAPLLLLRNSDCRPVAVPRCT
```

Alignment Scores:	
Pred. No.:	8,268-11
Score:	306.50
Percent Similarity:	38.17%
Best Local Similarity:	23.33%
Query Match:	9.66%
DB:	1
Length:	10074
Matches:	140
Conservative:	89
Mismatches:	232
Indels:	139
Gaps:	25

US-09-857-669-2 (1-609) x AE003957 (1-10074)

Qy	82	ValLeuAspLysGluClnThrGlyPhe-----LeuAlaGluGluAlaProAsp	97
Db	1286	ITGTTGGTGAAGGACAGCGGCAATTTGGCTGGGAATATTTGCTCAGCCAACGCGATTC	1345
Qy	98	AsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu-----	115
Db	1346	CAGACCCGCCAACCCCTTGAGGCCCTTGCTATTATGCTCCACCATTGCGATTGATGCC	1405
Qy	116	---ThrGluLysAspGly---AlaTyrThrValHisIleThrProGlyProArgThrLys	133
Db	1406	CCGCGCCAGAATGATCACATCACCTGGTATTATTACGTGGACAAGGAGAGCGGTCGCG	1465
Qy	134	IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu	153
Db	1466	GTACGTACGGCGCAGCTGCGATGACAGAGCGCGTGGCCAGGAC-----CAT	1513
Qy	154	TyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAsp	173
Db	1514	TATCTTCAGCGGTGACCTGGAAGATTTCAGCGCTAAGCTTGGCGAAATATTCAACCATCCC	1573
Qy	174	SerTrpClnAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu	193
Db	1574	ACGTATGAAGCCACGAAAGTTCCGATCAACCGCGCGCTTGGCTGAGCGAGGCTATTTCCGAT	1633
Qy	194	AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn	213
Db	1634	GCCGATTTCAACCATCGTCGGGTGCGAGTGACCCGTGGCGCGCATCGACGACAGATTGAT	1693
Qy	214	ValValValAspSerGlyArg-----ProIleAlaPheGly---AspPhe	227
Db	1694	TTGATCTGGGAAGCGGGCGCGCTATGACATGGGTCCGCTTCGCTTCATATAGCATAT	1753
Qy	228	GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln	247
Db	1754	-----TTCCACGNAAGGGCTATTTATCCACTGGTGTACTGGGAC	1792
Qy	248	ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn	267
Db	1793	GAAGGTAGCTATTTCACCAAGGCAAGCTGGATCGCTGCGTGAGTCCTCACCAAAATTG	1852
Qy	268	GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---Arg	286
Db	1853	GACTATTTTCAGCAGATTGATATCCAAACCCAGCTGAGGAAGCGGACCCTGAAGGCAAT	1912
Qy	287	ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg	306
Db	1913	GTTCGGCTGATGCTCAAGCTCGAAGCTGCCAAAGCAAAATCTACACGGCGGCATCAGT	1972
Qy	307	LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn	326
Db	1973	TATGGCAGCGAAAGTGGTCCCGGT-----	1996
Qy	327	LysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAla	346
Db	1997	-----CTCGGTGCT	2005
Qy	347	GlyIleSerGlnPro---ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyr	365
Db	2006	GGTGTGGAGCGGCGTACATGAATGCGGTGACACAAAATGAATGCACGATTCGACAT	2065

Qy	366	AsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyLe-	-----	382
Db	2066	-----GCCCAAGTCTCAAAGCCCTCACCGGGCTATCAAATCCCGCGTTC		2113
Qy	383	-----TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAla		397
Db	2114	AAATGGGTGGTGGGTACATA		2137
Qy	398	GluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGly-		414
Db	2138	---TTCCTCCGACGTCCTATGATGAGCAGACCAAGTACATCGCTTCGCTAACGTGAAG		2194
Qy	415	-----AsnSerHisAlaThr-	-----	419
Db	2195	TTATCGCGGCACGTAGCGCCAAATCAACCGACATTTGACGCGCCACCGCTCACTGAAC		2254
Qy	420	MetLeuThrAlaSerTrpLysArgGlnLeu-	-----	429
Db	2255	CGCGTACGGGGGGTGGCGCTTATAGGCGCGGATGACGCGCACCACATACAGTGGCTTACCAG		2314
Qy	430	LeuAsnAsnValLeuHisProGlu---AsnGlyHisTyrLeuAspGlyLysIleGlyThr		448
Db	2315	CAGTCCACGCTGGTCTACCCCAAAATGGAGACCGACTATGTGATGTTCAGCATGCCACC		2374
Qy	449	ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAla-	-----	466
Db	2375	TTCCCCCGCAATGGAGCGCGCCACAGTGCCTCGGTGGTGGCGCAGTGCCTTCGGT		2434
Qy	467	---GlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe-	-----	480
Db	2435	TCCAAAAGTAACCTTACCCAGCTTCATGGCCAACTCGCTGGTTCATGGATTTGGGGGCC		2494
Qy	481	-----IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAsp		496
Db	2495	AGCAGCGCTCATCTTACGTGAGAGGCGGTACCCACTGGCGCAGCATTTGGTCGCA		2554
Qy	497	ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu		516
Db	2555	ATGCGCGCAACCTTGCCTTCTTGTCTGGAGCGGTAAATAGCATTCGTGGTACGCTTC		2614
Qy	517	AspSerIleGly-----LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeu		534
Db	2615	CGCGAGTGTGTCACCGTACTGCCAAACGGATGCTTCGCACCTCGGTGCCAAACATGTT		2674
Qy	535	LeuValGlySerLeuGluTyrGlnLeu-	-----	549
Db	2675	TTCAGTGTGCGCGTGAATTTTGATCTACTACAAAGGGGGCGCTTT-		2722
Qy	550	SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu		569
Db	2723	GGTGGTCCGTGTTCTGGGATACCGGTAGTGGC-----TTCAATCGCTATCCGGAT		2773
Qy	570	LysHis---GlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPhe		588
Db	2774	TGGCATACCGCGCTGGTATCGCTTGCCTTACCGTCCCGCTAGTCCGCTACGTGTG		2833
Qy	589	AspIleAlaTyrGly-----HisSerAspLysLysIleArgTrpHisIleSerLeuGly		606
Db	2834	GATATTGCGGTGGCTTGAACGCCCGGATTCAGATTTCAAAGTTTCAACTGTATCGACATTGGG		2893
RESULT	28			
LOCUS	AE009623/c	10029 bp	DNA	linear
DEFINITION	Brucella melitensis strain 16M chromosome I, section 180 of 195			
ACCESSION	AE009623			
VERSION	AE009623.1			
KEYWORDS	GI:17983938			
SOURCE	Brucella melitensis.			
ORGANISM	Brucella melitensis			
REFERENCE	1 (bases 1 to 10029)			

ORGANISM	REFERENCE
<i>Brucella melitensis</i>	1 (bases 1 to 10029)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group	
Brucellaceae; Brucella.	

AUTHORS

DelVecchio,V.G., Kapatral,V., Redkar,R.J., Patra,G., Mujer,C.,
 Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A., Lykidis,A.,
 Reznik,G., Jablonski,L., Larsen,N., D'Souza,M., Bernal,A.,
 Mazur,M., Goltzman,E., Selkov,E., Elzer,P.H., Hagius,S.,
 O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyridides,N. and
 Overbeek,R.
 The genome sequence of the facultative intracellular pathogen
 Brucella melitensis
 Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
 11756688
 2 (bases 1 to 10029)
 DelVecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
 Direct Submission
 Submitted (13-NOV-2001) Institute of Molecular Biology and
 Medicine, University of Scranton, Scranton, PA 18510, USA
 3 (bases 1 to 10029)
 Elzer,P.H. and Hagius,S.
 Direct Submission
 Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
 Center, 111 Dalrymple Building, Baton Rouge, LA 70803, USA
 4 (bases 1 to 10029)
 Kapatral,V., Los,T., Ivanova,N., Anderson,I., Bhattacharyya,A.,
 Lykidis,A., Reznik,G., Jablonski,L., Larsen,N., D'Souza,M.,
 Bernal,A., Mazur,M., Goltzman,E., Selkov,E., Haselkorn,R.,
 Kyridides,N. and Overbeek,R.
 Direct Submission
 Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
 Park Drive, IL 60612, USA
 5 (bases 1 to 10029)
 Letesson,J.-J.
 Direct Submission
 Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
 Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
 61 rue de Bruxelles, Namur 5000, Belgium
 6 (bases 1 to 10029)
 O'Callaghan,D.
 Direct Submission
 Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
 Kennedy, Nimes 30900, France
 Location/Qualifiers

FEATURES

source
 1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"
 /translation="MAFEIFGIHLGKKDDPDIIIDPKTYSVDVTTTGRKNADGKE
 ADLKSVEAGSLVSDADKPSAGSLAKARGDYRRILSALYGERGYGTISIKYDG
 REANDIPDTEIPENNAKVAITVDPPQFLFESRTAISNIAPPNGNRKQVTPPEAGFA
 PGEASGTLIKERLAEVAHREGYAKARVTDGVVADHADNRNSADIALDPEKRAY
 IGPVSVGTAMPDQFVAMTIGLKPQGEYDPPDIIENAKRLGRMEVFRAMTFEADKI
 EPGSLPITLNVQERKPRFPGAESTIDFGVTSYMHRLNLLGRGERLRFADKVS
 IGQSQNSDFPKNTYLLGASFAKPGVYPTDIDFVATLDKREVLDAYTETSINAKTG
 FTOIFSDLSGALYANASQHFVDVFGKRDFTAGLEGSLLYDSRNNKPPDSSGFYL
 VQNPQPEYFHNFAETRTAEGRTVHGFQGTDRVVLACRLKVGSLVGSIALDPPSQ
 LFLAGGGSVRGYYGNIGVSAGNHIIGRSIVANGEVTRITISIGAVAVDAGY
 VGEKSPDFSEQMRVGVGGLRYLTSLGPIRLDVAVPLNRRSGDPNPGFYVVGQAF"
 complement(2475..2990)
 /gene="BMEI1896"
 complement(2475..2990)
 /gene="BMEI1896"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical Membrane Spanning Protein"
 /protein_id="AAL53077.1"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"
 /db_xref="GI:17983939"

gene

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"

CDS

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"

product

1..10029
 /organism="Brucella melitensis"
 /strain="16M"
 /db_xref="taxon:29459"
 /chromosome="I"
 complement(71..1942)
 /gene="BMEI1895"
 complement(71..1942)
 /gene="BMEI1895"
 /codon_start=1
 /transl_table=11
 /product="OUTER MEMBRANE PROTEIN"
 /protein_id="AAL53076.1"

```

/protein_id="AAL53082.1"
/db_xref="GI:17983945"
/translation="MRMERQMKHGTPTITVSFAVLALSFLACGCONVLLDPKGVG
LEERDLILFATGIMLLVYVPIATLFFFAKYRATNEKADYQPDWHHSKIELAVMLI
PMALIVVGLTWTISHLKDPYPLESNKVPINVEVVALDKMKWLFYIPDGIATVNM
AMVNPVPNFKLTSNLMNSFFIPQLGSQVYTPMSQTKLHLJANQEGDPAGISANY
GSGFAQRFKAHAFDAQGDFKWDYAKAAGKEDLDRAFYASLVQPSKHAQVFN
YGDKALFHDIVNRCWDGKSVCLDENHROQMIREARASLRKSSPIDFSQWANDICTV
OPRLSOLN"
gene 8234..9199
/genes="BMEI1902"
CDS 8234..9199
/genes="BMEI1902"
/codon_start=1
/transl_table=11
/product="molybdopter in biosynthesis enzyme"
/protein_id="AAL53083.1"
/db_xref="GI:17983946"
/translation="MIFGELPLDRAEDATILAHTINIDALSRLKGTVLARSIDIAALRQA
GIDRVLAARLESIGDGEDAAALAILAGLLAGNTEPGPATIGRYNLASPDGVFCTDTA
GIDAINHDPRISAVATIRNHARVEAGOMVATIKIIPFAPGVLIETRAVHGRPLLO
VEKFRNQGLIOSRLPSIRETVLDRTELMLHKRVTRNGGAVVQTRVAHHDALATA
IAKVGACDILYFSAVADEADIIPRAIRRAGGETLRIGMPVDPGNLVLVGMGRK
YIIAAPSARSARENSLDRLMAGMELDATHLARMGVGGLL"
gene 9244..9852
/genes="BMEI1903"
CDS 9244..9852
/genes="BMEI1903"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C-552"
/protein_id="AAL53084.1"
/db_xref="GI:17983947"
/translation="MNSRNNKFIIMAFATVVFVMTGILSDLSLPSAPEKPGFAIE
VAGSGGESAAPPAKEEVSIALLOSADPARGEVFKQCACHTGDKGANKVGPHL
WGVGPRAASVEGFSYSAAMKEFGAAGNWDFFHLNKLFTSPKGFINGTAMGAGDKK
DNERADILYLRFTLSNPVPLPAADAAPADSDRKPAAEAPAN"
BASE COUNT 2368 a 2856 c 2881 g 1924 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.12e-10 Length: 10029
Score: 304.50 Matches: 154
Percent Similarity: 37.86% Conservative: 111
Best Local Similarity: 22.00% Mismatches: 279
Query Match: 9.60% Indels: 157
DB: 1 Gaps: 29

```

```

US-09-857-669-2 (1-609) x AE009623 (1-10029)
Qy 7 AlaLeuLeuProAlaLeuPhePheProHisAlaTyAlaProAlaAlaAspLeu 26
Dy 1993 GCCATGCTGTTCGGCGGCATATTTTCCTCCTTCGGCTGTCGCT- 1946
Qy 27 SerGluAsnTysAlaAlaGlyPheAlaLeuPhe- 10029
Dy 1945 -----GCATTGGCGTTCAGATTTTGGCATTCACCTATCGGGCAAGGACAA 1898
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIle- 57
Dy 1897 AAGCAAGACCCCGACATCGATCCCAAG---ACCTATAGCGTCGATGCCACCACT 1841
Qy 58 -----AspThrGluAspSerGluIleLysAspMetValGluGluHis 71
Dy 1840 GCGCATCGCAAGATGCGGACCGCAAGAGCGGACCTGGAATCAGTATCGAAGGCGCA 1781
Qy 72 LeuProLeuIleThrGlnGlnGluValLeuAspTysGluGlnThrGlyPheLeu 91
Dy 1780 TCCGGTCTGGTTTCGGATGCGGACAAAGCCGCT-----TCCGGCTCGCGGGTCTGCTG 1727
Qy 92 AlaGluGluAlaProAsp-----AsnValLysThrMetLeuArgSerLysGlyTyPhe 109
Dy 1726 GCCAAGGCGCGCGGATATTCGCGGTATCCTGTCGCGCTTTATGTGTGAAGCCGCTAC 1667

```

```

Qy 110 SerSerLysValSerLeu-----Thr 116
Dy 1666 GCGCGCACCATCTCCATCAAGTGCATGGCGCGAGGGAACGATATTCGCCCGCACACG 1607
Qy 117 Glu-----LysAspGlyAlaTyThrValHisIleThrProGlyPro----- 130
Dy 1606 GAAATTCCTCCAAATATCGAAGTTTGCATACGCTTGCATCGGGCGCGCAATTCCTGTTC 1547
Qy 131 ---ArgThrLysIleAlaAsnVal----- 137
Dy 1546 TCGCGCACGGCTATTTCAAATATCGCGCGCGCGCGCAACAGCGCGGATAAAGTGCAG 1487
Qy 138 -----GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Dy 1486 ACGCGGAAAGAGCGGCTTTTGCA---CCGGGCGAGGAGCAAGTCCGGCACCATCTTCG 1430
Qy 153 GluTyTyTrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGln 172
Dy 1429 AAGCGGAACGCCCTTCGCTGGAAGCCTGGGTGAG----- 1394
Qy 173 AspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyPro 192
Dy 1393 -----GAAGGATATGCC 1382
Qy 193 LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeu 212
Dy 1381 AAGCGCGTGTACGCGGAGGAGTGTCTGCCGACCATCCGATACCGCGTTTCCGCC 1322
Qy 213 AsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Dy 1321 GATATGCACCTCGATCCGGGTAGAAAGCCTATTATGGCGCGTCAGCTGGTGGGAC 1262
Qy 233 GlnArgTyProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTy 252
Dy 1261 GCGCGCATGGAACCGCAATTCGTGCTGATGACGGCGCTGGAAGCGCGTTCAGGAATAT 1202
Qy 253 AspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTySer--- 271
Dy 1201 GATCCCGAC-----GATATTGAAATGCCAAGAGCGCTTGGCGCGCATGGAAGTC 1151
Qy 272 ---GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---ArgValProVal 289
Dy 1150 TTCCCGCGCATGACCTTCGAGGAGCGCAAGATCGAACCGGATGCGAGTCTCCGATA 1091
Qy 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSer 309
Dy 1090 ACGCTCAATGTTACAGGAGCGCAAGCGCGCTTGGCTTCGCGCGTGAATATTCAC 1031
Qy 310 GluTyTyTrGlyLeuGlyLysIleAlaTyTrAspTyTrAsnLeuPheAsnLysGlyTy 329
Dy 1030 ATCGATGGCTTCGGTGTGACGAGCTACTGGATGCGACCGCAACCTGTTGGCGCGGCG 971
Qy 330 -----IleGlySerVal-----ValTrpAsp 336
Dy 970 GCGCTCGCTTCGATGCGAAGTTCAGCGGATTCGCGCTCCAGGATAATTCCTTCGAT 911
Qy 337 MetAspLysTyTrGluThrThrLeuAlaAlaGlyLysSerGlnProArgAsnTyArgGly 356
Dy 910 CCCAAGAATTATACCTATCTTTCGGTGCCTCTTCGCCAAGCGCGCGCTATACGCCG 851
Qy 357 AsnTyTrThrSerAsnValSerThrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Dy 850 GAC-----ACGGATTTTGTGCG--CACACTGATGCCAAGCGGAGTGTGGAGCG-- 801
Qy 377 AlaPheSerGlyGlyIleTyTrValArgAspArg-----AlaGlyIleAspAlaArg 394
Dy 800 ---CTATACGGAACCGTCCATCAATGCCAAGACAGCGGTTTACGCAGATTTTTCAGCG 744
Qy 395 LeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlu 414
Dy 743 GTTCTCGGCGCGCTCTATGCCAATGCAAG-----CGAGGCGCACTTCGTTGACGATGT 690

```

```

Qy 414 yAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu----- 430
Db 689 GTTTGGCAAGCGGATTTCACGACTCGCGGCTTGAGGCAATTCGTCTATCATAGCCG 630
Qy 431 -AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyIleValThr----- 448
Db 629 CAACAATAAGCCGACCAAGTTCCGGGTTCTATCTGTGGCGCAATATCCAGCCCTCTA 570
Qy 449 -----ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleAArgThrSerAlaArgAl 466
Db 569 TGAATCCATATGCGCAATTCGCCACCGCG----- 539
Qy 466 aGlyTyrPhePheThrProGluAsnTyrLysLeuGlyThrPhe----- 480
Db 538 -----TTCACCGCTGAAGCGCGCACCTATCATCGGTTTCGCCAGACGGATCGTGT 489
Qy 481 -IleIleArgGly-----GlnAlaGlyTyrThrValAlaArgAspAsnAlaaspValPr 498
Db 488 GGTTCGCGACGCGCGGTGAAGCGCGCTCCATCGTTGGCGGCTCGATTGCAGATCTGCC 429
Qy 498 oSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuaspSe 518
Db 428 ACCAGCCAGCTTTCCTCGCTGGCGCGCGGCTCGGTTTCGGGCTACGCTATCCCAA 369
Qy 518 rIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySe 538
Db 368 TATCGCGCTCAGCGCGCGAAGCGGACATCATCGCGCGCGCTCGCTCGCGAGGCCAA 309
Qy 538 rLeuGlyTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisaspMetGl 558
Db 308 TGGCGAAGTCGCGACACGCGATTACCGATTCCATCGGCGCGGTTCGCTGGATCGCGG 249
Qy 558 y-----AspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGl 574
Db 248 CTATCGCGGTGAATAATCTTCCCGGATTCTCCGAG---CAGATCGCGCTCGGGTGGG 192
Qy 574 yLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
Db 191 CGCGCGCTTCGCTATCTTACAGCCCTTGGACCGATCGCGCTTGATGTCGCGCTCCGCT 132
Qy 593 -----GlyHisSerAspLysIleArgTrpHisLeSerLeuGlyThrArgPhe 609
Db 131 CAACCGCGCTCGCGCGATCCGAACACTATGTTTCTACGTAGGCATAGACAGCGGCTTT 74

RESULT 29
AF088896/c AF088896 39013 bp DNA linear BCT 25-MAR-1999
LOCUS Zymomonas mobilis fosmid clone 42C11, complete sequence.
DEFINITION Zymomonas mobilis fosmid clone 42C11, complete sequence.
ACCESSION AF088896
VERSION AF088896.1 GI:4511972
KEYWORDS
SOURCE Zymomonas mobilis.
ORGANISM Zymomonas mobilis
Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
Zymomonas.
REFERENCE
1 (bases 1 to 39013)
AUTHORS Lee, H. J. and Kang, H. S.
TITLE Sequence analysis of 42C11 fosmid clone of zymomonas mobilis ZM4
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 39013)
AUTHORS Lee, H. J. and Kang, H. S.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1998) Microbiology, Seoul National University,
San 56-1 Shillimdong Kwanakgu, Seoul 151-742, Korea
FEATURES
source
Location/Qualifiers
1. 39013
/organism="Zymomonas mobilis"
/strain="ZM4"
/db_xref="taxon:542"
/clone="fosmid 42C11"
666..1940
/gene="sah"
666..1940
gene
CDS

```

```

/gene="sah"
/codon_start=1
/transl_table=11
/product="succinylarginine dihydrolase"
/protein_id="AAD21533.1"
/db_xref="GI:4511973"
/translation="MTIVSEVNFDELGLTPHNYAGLSRGVNASALHAGOPSYPRQAALQ
GLEKMKHMDMGLTQGVLPPLRPVTHLLHLGKGDGKTLKQAARMIGCVFNLC
ASMAANAATVISEFSDHGRVHFITANLATMLHRLEAOTTYAOLNOMFNSCFPA
MHPLPCGQHFSDGCAANMRITSAHRTGGINLFYVGEKNDIYPAOKIARSAOVARL
GEVKPLAWFIFOKREAIAGAFHNDVAVANEIVLLAHAEAFEDGGEWIKRIAEKID
GIPIIIDNITLEAVKSYLFSQIVTLKDRTMALIPQEVKSDPAWBTWNRIISGN
NPIKAVYVDVRESMANGGCPACRLRVLPLSKAALEAVDQREFLTTPKWEKLYQLVEN
FWPEKITPDOLVLPWLKTAVRHAWALTSWLG"
complement(1937..3067)
/notes="hypothetical protein; orf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21540.1"
/db_xref="GI:4511980"
/translation="MIPLFOLMLMDIFDSKFFLDKWTGKACKGTLTAKGYASPDAA
SDYSALNITLDKATLARDEMKTSGNLDITSNRKDGSGIKGUKLPETRIFRFIMD
NKETVHDLGKQVEDTAHQTKESSASASMLASANKLDIRISAPQLCFVAGHL
ESEMADLRVGTATAPNITGDMVRITGYTFAGRRFIDHGEIQGTGNGPNPALDI
TAAATVDDITATVKIGGFANKPEITFSSPSLAODEILSRLLFGSSVTSLSVAQVQL
AAALNTLRTGGKGFDPDLKLSVIGIDRLRVVCANSTGGTSLAAGKYLFKNVYMEV
YDTHGFTATQIRISLTRLSLLSEASSFGSSNVSLRYSKDI"
complement(3317..6106)
/notes="orf2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21536.1"
/db_xref="GI:4511976"
/translation="MTDQNYNPERQNSSDSIKEKNPNVIAQPTESANKPNTSPFON
LKHLMGCVFTGIALFILALILFDLTSTGHLIVRYINKYTESGLVKIKRIDGS
IYQAFRLVDIKAYDQRFELADPSEVWRPLDYGRIITIKLHAPRIEVLNPELK
PYKSNPDALLPDIHLIDQIHLDRIDISKNLDQONHRLHIHGLRALIKGQITVKAFA
GADKQDISGGDTLSLQDAKPTANLLVVEAHLHAPVGVVDGLTRGLRPLTMDTAGK
GNQDWRGYIQAGSGDKLUDMOIHNNKDSVHLDGAVYPSLLPDPDPPQAPFKKGLQ
IDLTHNKNVNDIRGATGEMFRLTAGVDSKSSFNMSYDLAIPNSQKFLGDDLI
SPKNTFSSVLSGLVSPNIDMLRKAGQGWQEKIAENTISQSGKMLVDGDSLMPGL
SYGKISGLDPLRLNLENLKIIEGAINYGDHVLVDHSNISRYIHANTTITDLYTKK
WRAAVKGDIPYKIDGILSLTIDSHLOPEKNGLGVMGEMMKGRDQKPKALARAF
GGAVALTRMIANANIRMPRESLASPEHLECGNGYQPKGNINLEVRGKSKSYGAM
TLDVGIVSHPVITAKALGTVNGNLAQTKSNLYHQFELNGRLGDIIRLITEEQATQ
AVNADTAIDKAHFVYQPIDIESGRLRGKAVFTDPAFOEALLNNIRYNDLFLKQAA
GRAFYQKEQEGGLTQGNLPAAGASVPFNVAIHSFSADVIRANAQGVNHLNWHM
AAPVEVRSGKNYDISKIVTLQPKGHIDLSGHYGQNSVAALOADIADLSEANLLWPE
INLSGOLSAVANADLTGNIPQISARLALNNFTRSVTGLSQPVNLVVAATNDKGLVS
EAVIROKQTMIGQIR"
complement(6106..8322)
/notes="hypothetical protein; orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21534.1"
/db_xref="GI:4511974"
/translation="MLFRHYFRQIHRICILSVGLISLSHAEDVSTPKTDSGNKAASP
NPAIDPDPSPFFDHLTQSPTEASKADGSSHLPLRTSESSQGSNGRSGVKALDT
ESDLDKYNLLANSDESEKTAQNIETTSVTLDPALSEPLDLSHPMDQNEHVLAK
ETDLYKVDYKGLSNIDTQAFKNYSVLLSNKADSLSVIGARATNDQELNRLILR
SQYTDCKAALSITPIKEGQINRYDVNAGSYKLGQINTLGSEEPRLRVARVALMLH
RGPIITATHIEQAENNLITLPHYGFPAKIGDRTLLDDETHGTDTLTPVNAAGVTS
YGSIIYSNNKHIVDAKHSIARFEQGYDQSDMWDLQALAASTLSFHSVSPETFA
TGRKREDGSEIVDLVRQGRKKHSTAVTGGTGGEGKAGQSWISRNYPPEGSULTF
SGILGTQQOOLSALFNRRNAGARDRVQIGLTAGRERYDAYNGYSFSLGSSLSQSTQ
LWQKRWTYSIGALETOTNERSYDFSERLTRYLIAALPGOLGYDRSNLMNPTKGY
RLNMRFPETSIGSGURGVYRMLFDASGYYPVADNVYLAQFRVASTEGASVQELAPG
RRIYAGGSGSVRGYQYQGLGPKDPYKDAVGGLSEELAFEVRFRYFNGFLPFDIAPG
VYEDAITPHNLRFGTGICARYYTATGPFPRIDLATPTARQGESRISVYISIGQAF"
complement(8322..9662)

```

gene

```
/gene="cpxA"  
complement(8322..9662)  
/gene="cpxA"  
/codon_start=1  
/transl_table=11  
/product="chemotaxis protein"  
/protein_id="AAD21535.1"  
/db_xref="GI:4511975"  
/translation="MLELLLSLFSEASQLTYYADATQAITERVLVARRIISODRP  
SRPMIANFNSREIRFEFMSRDIKHINSDSLVDLQOGLSHSDSOILPMDIHID  
LAPKGRDHLGQIKLSDGSWLYFQTLVNPAPNPWLNRLTITAILVIGAGLVGLLI  
QAFVDPURSLKAVKQDGKHIFTETGPKELRHVQAQFNQOMRLEHMLRRETEML  
AASHVDITPLARLOLRASFEEDIDREAVADLQEMATMIKSVLPQFGDQDEVRK  
LTDIVAIETLVNAADKAGDISYEGSPSHDVTTRPLLMKRAISNLJENALHTYGKAI  
VRFHENGVIQVDDAGPGIPEDQIEQVLQPPRLDTARQNTAGLGLATGTRTV  
KREGKLSLNLQEGGLRAEITFLPAAQNKIRKKGKGYFSARKGLELLRRPK"  
complement(9737..10462)  
/gene="arC"  
complement(9737..10462)  
/gene="arC"  
/codon_start=1  
/transl_table=11  
/product="aerobic respiration control protein"  
/protein_id="AAD21537.1"  
/db_xref="GI:4511977"  
/translation="MASOTIMVEDDASRLARLAVRLKENCYIPCPVSGSEEMWEEIQ  
SRVVDLMDVMLPGKNGLDLTRLSNDSYMAIPVIFLSARODETRVVGLGGLD  
YIPKPCCKELLVARVAVLRQAVKEADNPQSHLIHFEGWLDPRRLHSPGTAG  
VELSGAEFDLLSFLOHAQVIGRERLLELSRSLSEASDRSDLVLSRLRLKLSOT  
EAGMDIRITKGVTMTVPVVSQD"  
complement(10491..10865)  
/gene="hise"  
complement(10491..10865)  
/gene="hise"  
/codon_start=1  
/transl_table=11  
/product="phosphoribosyl c-AMP cyclohydrolase"  
/protein_id="AAD21538.1"  
/db_xref="GI:4511978"  
/translation="MONDQYGNKLNPAWDEKGLITAVLTDAKTGLLVMVHNKESF  
ELSMATKEAFWMSRKKIWRKGEESHVWKIQEIRIDCDODTLWLKVIWPGPACHTG  
AQSCFYRLDEGHLVTIPKDOI"  
complement(10920..12113)  
/gene="ygcA"  
complement(10920..12113)  
/gene="ygcA"  
/codon_start=1  
/transl_table=11  
/product="RNA methyltransferase"  
/protein_id="AAD21539.1"  
/db_xref="GI:4511979"  
/translation="MKIIRLASKGQGVTEGQFVPSVDDYISDDGKLEFGAHTE  
PVCFHESVCGGCRLOYADEVYKNFLDRIAFAFHQAALSAPLYLKTALHSPVSRVR  
ALRAFRAKKTLCYKNTSSHLVDIVCECPDLIDNLFKAAMDLSRFQKWLAPRSQAQ  
IEMTADQGLDCLVMPFPETLEATEITAFPAEKGFARLSIDQGVGVETRWSESERV  
VTGLGAVPTLPAHFQATQDGEOTLVHMYKEAVGDAHFVADLFSGLTALSFQKDK  
RVTAAGMDRAVLAKQAALAKAVVEHRLFRRLQKDELVRPFCIIILDPRAGA  
KEOTIANLILPNCRIYVSCNPATFARDAKTLLAEGVHLHWKPVGQFVSLHVEWG  
LFTKM"  
complement(12110..13564)  
/note="hypothetical protein; orf4"  
/codon_start=1  
/transl_table=11  
/product="unknown"  
/protein_id="AAD21559.1"  
/db_xref="GI:4511999"  
/translation="MTISKLETLPQHDSRSTRSALNRPRLRLQAMRAADMAAKKCS  
SVDDMEKAKVAASITRSPAGNPALILCGPNNNGDGYVAVRYLAEMGIPYSVAY  
SEPTDAAKKARSLFGKVTSLDEAEAPLLIDALFCVGRSRLSDNALLSKRLTALSRA  
ASRLAIDVPGVDTKGSLSDIPADFLTIAIGALKPAHRLPSATICCRVAGDID  
LITSNPDLEIPRLAAPTADYKYSRGVVAITACAMAGASLLSAVAARACAGGYA  
LTGATLSSGPMALVHRSNRSREATERINALVIGPLGRDELGVQRUSMALDTSHPLV  
LDADALFLLSREGDLURNIOQTITITPHEGEPKRLFEIGSGDKLITRVTAKAALSGLV
```

```
VVLKGDVTIATPDGRAAIISQPAVSWLANAGTGDVLSGICGAMLARGMDFDAACTAV  
WLHGEAARLAGPGMIADDLTHILRSAPVEVTVAS"  
complement(13922..14641)  
/note="hypothetical protein; orf5"  
/codon_start=1  
/transl_table=11  
/product="unknown"  
/protein_id="AAD21560.1"  
/db_xref="GI:4512000"  
/translation="MISLGNSSLLIIVDHASAFVPEDIDLGIDKRLYDEHVAVDKGV  
QTLEQCUCQACNGVFAAOSRLVLDNRDENDFLGLMPAISDGYIHTGNTGLTADK  
EARLSRYPYHQHIDAIQKQRLIISLHSTPCIKSOLNCRYPWQIGLLYNPRVSGS  
ATQATIPLLQAHLVGENEPYSGRIILNTTMRHAENRPIYLLFEIRQDMLEBEANTI  
GNWCAILUKVINETLISLKNK"  
complement(14821..15606)  
/note="hypothetical protein; orf6"  
/codon_start=1  
/transl_table=11  
/product="unknown"  
/protein_id="AAD21541.1"  
/db_xref="GI:4511981"  
/translation="MPNGYHALETIFAFAKDGDILOAEANDTEENLTITGPFSEGLEA  
NKONLVRAVTAURQACNPKIIPAGFSIILDKRLPVAAGIGGGSADAAMLRMIQGHYO  
IPHELILSLANSIGADVPACVDSCLVRGEGVEKLTQIGDRSLERKPLLLVNPVSGS  
TPWIFKWDQVDRGALDSGSIILGAARSGRNDLEPPARKILPIIIEVVEVWLOQKQKVS  
FSRMGSGATCFALFDEIEDORTAYKKLNIIDHPEWALSLLR"  
complement(15964..17607)
```

Alignment Scores:
Pred. No.: 1.34e-09 Length: 39013
Score: 300.00 Matches: 168
Percent Similarity: 37.50% Conservative: 99
Best Local Similarity: 23.60% Mismatches: 277
Query Match: 9.45% Indels: 169
DB: 1 Gaps: 33

US-09-857-669-2 (1-609) x AF088896 (1-39013)

Qy 1 MetMetileLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 8048 CTAATGGCAGAGGGTTAAACGCTTTAGACACCGAACTGTTTCA-GATCCGAGAACCTC 7990
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 7989 TTGCTAATTCTGCGCATCAGAGAAAACGGCA-----CAGAAT 7948
Qy 41 SerProaspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 7947 ATCGAAGAGACAACTCTCTGTTACGCTTGATCCTGATCCTATCAGAACCCACTGCCTGACCTT 7888
Qy 61 AspSerGluileLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 7887 TCTCATTATTGCTGATATG-ATGAACATGTGGTCTTGCCCAAGAAGTGACGAC 7831
Qy 81 -----GluValLeuAspLysGluGlnThr----- 88
Db 7830 CTCGGATACAAGGTAGATGTCAAGGCTCTCAGCAATATCGCATACTGATGACGCTTCAAA 7771
Qy 89 -----GlyPhe 90
Db 7770 AACTATTCTGTTTACTATCGAATAAGACCGCATGACCTATCGCTTATTGGTCT 7711
Qy 91 LeuAlaGluGluAlaProaspAsnValLysThrMetLeuArgSerLysGlyTyrPheSer 110
Db 7710 AGACAACCAATGACCAAGACTTGATTAAACCGGATTTTACCATCTCAAGGCTATATGAT 7651
Qy 111 SerLysValSerLeuThr-----GlyLysAspGlyAlaTyrThrValHisIle 126
Db 7650 GGCAAGGCAGCTCTTCTTATTATCCGATAGAAAAA---GGCAGTATAAATGTTCCGCTAT 7594
Qy 127 ThrProGlyProArgThrLysIleAlaAsnValGly----ValAlaIleLeuGlyAspIle 145
Db 7593 GAT-----GTCAATCGAGTTTCGATCTACAGCTAGGACAAATA 7555

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 7554 AATTGACCGGAGTCTAGGAGGAACTTAAGAAATTGCCCGCTCGCATTAATTTGCAT 7495
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 7494 AGAGGGATCCCGATTATGCGACCCATATAGACGAGCTGAAATAATATCTCAAACT 7435
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 7434 CTCCCGATTATGTTATCTTTGCAAAATTTGGCGATCGGACAATACGCTGGATGAT 7375
QY 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 7374 GAAACCATACAGGTGACTATACATACCGTAATGCAGGAATGTTGGATCTTACGGT 7315
QY 226 AspPheGluIleThrGlyThrGlnArg-----TyrProGluGlnIleValSerGlyLeu 243
Db 7314 TCTATTATCGTTAGCAACAATAAGCATATGTTCTGGATGCCAACACATCTCTCATATT 7255
QY 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAla 263
Db 7254 GCCTGTTTAAAGAGTCAAGCTATGATAGCCAGATGCTGATGATCTCCGCAAGCT 7195
QY 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln----- 276
Db 7194 CTAGCGCAACAGTCTATTTCTCATGTTTCTAGTTGACCCATTCGAACGCGCGNAAA 7135
QY 277 -----AlaAspPheAspArgLeuGlnGlyAspArgValProVal 289
Db 7134 CGGGAAGATGGAAGCAATGTTGGATCTCGAGTTAGGCGGGA----- 7090
QY 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSer 309
Db 7089 -----CGAGGAAAGAACAC-----AGCATGCTGTGTACGGGT 7057
QY 310 GluTyrGlyLeu-----GlyGlyLysIleAlaTyrAspTyrTyr 322
Db 7056 GGTATGACCGGCTGAGGCTTTAAGCGCAAGGTAGCTGATCAGCCGTATATTATTC 6997
QY 322 ----- 322
Db 6996 CCACCGGAGGATCAATACTTTTGGGGATTTGGTACCGCCAGCAACAGCTTCT 6937
QY 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys---TyrGlu 341
Db 6936 GCTCTCTTTAACCGCAATAATGCTGGGCA-----AGAGACCGTGTATTTCAA 6889
QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 6888 ATGTTGTTGACCGGGTGGGAGAGATACGACGCTTATATGTTATAGCTTTCTCTC 6829
QY 362 AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly 381
Db 6828 GGAGGGAGCTTGTGCGCAATCTACCCAGTTATGGCAAAAGCC----- 6784
QY 382 IleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAla 401
Db 6783 ---TGG-----ACTTATTCTATCGGAGCAAGAAATTAACACAG 6751
QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr----- 419
Db 6750 ACCAACAGCGG-----AGTTATGATTCTCTCGATCCGAGAGGTTAACCCGT 6703
QY 420 ---MetLeuThrAlaSerTrpLysArgGlnLeu-----LeuAsnAsnValLeu 434
Db 6702 TTTTATCTGATCGCGGCTCCCGGCAATTTGGATATGACCGTTCACAACTCTGATG 6643
QY 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGly-----ThrThrLeuGlyThr 452
Db 6642 AATCAACCAAGAGTTACCGCTTGAATATGCTGTTTATGCTCTGAAACCTTCAATCGGTTCT 6583

QY 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro 472
Db 6582 GGGCTACGTGGGTATCTCCGAATGCTTTTGTATGCG---AGCGGTACTACCCGTAGCA 6526
QY 473 GluAsnLysLysLeu---GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAla 491
Db 6525 GATAATGTTGTGTGCTGCTCGGTTTTCGCTTCTCTATTGAAGGTGCCCTCGGTT--- 6469
QY 492 ArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerVal 511
Db 6468 -----CAGAATTGGCGGCTCTCTCGGATTTAT---GCAGTGGCGGCGCTCTGTT 6418
QY 512 ArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn-----GlySer 527
Db 6417 CGTGGCTATGTTATCAGCAGTTAGCCCCAAAGACCCATATAAGATGCCGTGGCGGA 6358
QY 528 ValLeuProGluArgAlaLeuValGlySerLeuGluTyrGln----- 542
Db 6357 CTGTCGGAAGAAGAG-----CTTGCTTTCGAAGTGGCTTACCGTTTGGTAAATTTGGT 6304
QY 543 ---LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 561
Db 6303 CTGTGCTCTTCAATGAT-----GCAGGCGAGTATATGAA-----GACGCTATC 6259
QY 562 AlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 6258 CCAACCTTCAT-----AATTTAAGATTGGTACTGTTGGTGGTCTCTTACTATACC 6205
QY 582 ProLeuAlaProPheSerPheAsp-----IleAlaTyrGlyHisSerAspLys 597
Db 6204 GCTTTTGGTCTTTCGCGATGATCTTGCACACCTATAGTCGCGGACCTGGAGAATCT 6145
QY 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 6144 CGTATTTCAGTATATATTTCGATCGGTGAGCTTTC 6109
RESULT 30
LOCUS AY010120 20310 bp DNA linear BCT 13-JUN-2002
DEFINITION Xanthomonas oryzae pv. oryzae xanthomonadin biosynthetic gene cluster, partial sequence.
ACCESSION AY010120
VERSION AY010120.2 GI:15055553
KEYWORDS Xanthomonas oryzae pv. oryzae.
SOURCE Xanthomonas oryzae pv. oryzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.
REFERENCE 1 (bases 1 to 20310)
AUTHORS Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.
TITLE Genetic locus encoding functions involved in biosynthesis and outer membrane localization of xanthomonadin in Xanthomonas oryzae pv. oryzae
JOURNAL J. Bacteriol. 184 (13), 3539-3548 (2002)
MEDLINE 22053219
PUBMED 12057948
REFERENCE 2 (bases 1 to 15118)
AUTHORS Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2000) C.C.M.B., Uppal Road, Hyderabad, A.P. 500 007, India
REFERENCE 3 (bases 1 to 20310)
AUTHORS Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) C.C.M.B., Habshiguda, Hyderabad, A.P. 500 007, India
REMARK Amino acid sequence update by submitter
AUTHORS Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2002) C.C.M.B., Habshiguda, Hyderabad, A.P. 500 007, India
REMARK Sequence update by submitter

Qy	266	GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu-----Gln	283
	:::	:::::	:::
Db	941	AAGCTGGATTACTTCAGCCACCATCGATATCCAGGCCAACCCGAAGAAGCCGACGACCAG	1000
		:::	:::
Qy	284	GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr	303
Db	1001	GGG---CGCGTGCGGTGCAGGTCAAACTCACCCGCCAAGCGCAGCGTGTACACCGCA	1057
Qy	304	GlyIleArgLeuAspSerGluTyrGlyLeuGly-----GlyLysIleAlaTyrAsnTyr	321
Db	1058	GGCTGAGTACCGCAGCAGAGACGCCGCCGGTGTGCCGGTGGCGGTGGACGGCGGTAC	1117
Qy	322	TyrAsnLeuPheAsnLysGlyTyrlleGlySerValValTrpAspMetAspLysTyrGlu	341
		:::	:::
Db	1118	GTCAATGCG-----CGTGGCCACAAGATGGACAGCAGCTGGATTACGCCCAACACCGC	1171
Qy	342	ThrThrLeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrp	359
	:::		
Db	1172	AAGAGCGCTACCAACAGCTACCGCATACCGCGCGTTCGGTTGGCTGGCTGGTACACC	1231

Pred. No.:	1.96e-09
Score:	292.00
Percent Similarity:	40.43%
Best Local Similarity:	25.12%
Query Match:	9.20%
DB:	1
Gaps:	
Indel:	
Mismatch:	
Conserve:	
Match:	
Length:	

Qy	58	AspThrGlnAspSerGluLeuLysAspMetValGluGluHisLeuProLeuLeuThrGln	77
Db	323	GAGGGCAGCACCGCAGATGATCGAGACATCCAGGCTCTTTGTCTCTCTAT	376
Qy	78	GlnGlnGluGluValLeuLysPlysGluGln-----ThrGlyPheLeuAlaGlu	93
Db	377	-----GAGGCTGTGGCAGGACACAGGGCGATCGCGTGTGAGTATCTGTGGCA	427
Qy	94	GluAlaProAspAsnValIysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal	113
Db	428	CAGSCCGAGCGCAGACGGCGAGGCGCTGGAGCCGTTTGGCTATTACTCGCAACCATC	487
Qy	114	SerLeuThr-----GluLysAspGlyAlaTyrThrValHisIleThr-----ProGly	129
Db	488	ACGCTGGCTGGCCACCGCGCGGACAAAGGTGACGCTGGTACGCTGGATCGTGA	547
Qy	130	ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly	149
Db	548	GAGCGGTGGGGTGGCCACTCGCATATTTTCGATCACCGGTGGGCCGAGCAGGAC	604
Qy	150	AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp	169
Db	605	-----CGTTATCTAGGGCAGGACCTCAAGCAGTTCCAGCGCGCGAGGGCAGGTC	655
Qy	170	PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys	189
Db	656	TTCAGTCATCCGAGTAGCAGCAGCCAGAGTCCGATCACCGGCGCCCTGGCCGAGCGT	715
Qy	190	GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr	209
Db	716	GGCTACTTCGATCGGACTTTACCAAGCGCGGTAGCGATCAGCGCGCCGAGCAGCGG	775
Qy	210	ValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGly-----	225
Db	776	GCCGATATCATCTGAATTGGGACACGCGCGCGCTACGACATGGCGAAGGTGGCGTTC	835
Qy	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuLeuArg	245
Db	836	GACTACGACTAC-----TTCGGCGATGGTCTGTTCACCGCTGGTGTAC	880
Qy	246	PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAlaLeuGlu	265
Db	881	TGGACGAGGGCAGCTACTACCAAGGACGCAAGCTGGATCGGCTACCGAGTGGCTCAC	940

US-09-857-669-2 (1-609) x AE012071 (1-10839)

Qy	58	AspThrGlnAspSerGluIleuLeuAspMetValGluGluHisLeuProLeuIleThrGln	77
Db	10386	GACGCGACGACGCGGAGATGTCGAGAACAATCCAGGTCCTCCCTGCTCTAT-----	10333
Qy	78	GlnGlnGluValLeuAspLysGluGln-----	10333
Db	10332	-----GAGCCGTGGCAAGCAGGACGAGCGAGTCGCTCTGGAAATATCTCTTGCCA	10288
Qy	94	GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal	113
Db	10281	CAGGCGACGCGGACGCGGTGGAGCGGTGGAGCGGTGGCTATTACTCGCCGACCAT	10222
Qy	114	SerLeuThr-----GluLysAspGlyAlaTyrThrValHisIleThr-----	129
Db	10221	ACGCTGGCGCGACACGCGCGCGACAAAGTGACCGTGGTGATCACGGTGGATCGTGGC	1016
Qy	130	ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly	149
Db	10161	GAGCGGTGGCGGTGGCCACCTCGCATATCTCCATCACCGCTGGCGGACGAGAC---	1010
Qy	150	AsnLeuAlaCysTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp	169
Db	10104	-----CGCTATCTGGCGCAGGATCTCAAGCAGTTCGAGCGCGCGAAGCGCAGGTC	1005
Qy	170	PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys	189
Db	10053	TTCAGCATCCGCGAGTACGACGACGAGGTCCGCATCACCCGGCGTCTGGCGAGCGT	9994
Qy	190	GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr	209
Db	9993	GGCTACTTCGATCCGCACTTCCACGAGCGCGCTGGCATCACGCGTCCGAGCATGCG	9934
Qy	210	ValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly-----	225
Db	9933	GCCGACATCATGTAATTTGGGACACGCGCGCCGCTACGACATGGCGAAGTCCGCTC	9874
Qy	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
Db	9873	GACTACGACTAC-----TTCCCGGATGGCTGTTCGACCCGCTGGTGAC	9829
Qy	246	PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu	265
Db	9828	TGGACGAAGCGACGCTACTACCAGGAAGCAAGCTTGACCGATTGCGCGAGTCGCTGAC	9769
Qy	266	GluAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu-----	283
Db	9768	AAGCTGGATTACTTCGACCACTACGATATCCAAACCCAGCCGAGAGCGGACGACGAC	9709
Qy	284	GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr	303
Db	9708	GGC---TGCCTGCGGTGGAGCTCAAACTCACCCGCGCAAGCGCACCGTCTACACCTCC	9652
Qy	304	GlyIleArgLeuAspSerGluTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyr	321
Db	9651	GGCTTGAGCTACGGCAGCAAGCGCGCGCGTGTGCGGGTGGCGTGGAGCGCGGTAC	9592
Qy	322	TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu	341
Db	9591	GTGAACGCG-----CGCGGCCAACAGATGGACACGACGCTGGATTACGCGCAACGCG	9538
Qy	342	ThrThrLeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrp	359
Db	9537	AAGAGCTGACCAACGATTACCGCTGCGCGGCTTCCGTTGGCTGGATGGCTGGTACACC	9478
Qy	360	ThrSer-----AsnValSerTyr-----AsnValSerTyr	365
Db	9477	GCCTCGCGCGGTGTACGACGACGACCGAGTACATCGACCTCGCGCAAGCTCAAACTC	9418
Qy	366	AsnArgSerThrThrGlnAsnLeuGluLysArg-----	376
Db	9417	ACCGGACGCGGACGCGGACGATCAACGAGCGCTGGAGCGCATCGCTCGATCAAGCGC	9358

Pred. No.:	9.28e-10	Length:	10839
Score:	291.50	Matches:	149
Percent Similarity:	37.32%	Conservative:	85
Best Local Similarity:	23.76%	Mismatches:	245
Query Match:	9.19%	Indels:	148
DB:	1	Gaps:	24

Qy 377 -----AlapheSerGlyGlyIleTrp----- 383
 Db 9357 TTGCGGAGCGTGGCTTTCAGCAGCGCGGATTCGAAGCGCGGTGTACGAGAC 9298
 Qy 384 -----TyrValArg----- 386
 Db 9297 TCCACACTGATCTATCCCAACTGCAGGCCAACTACGTCAATGTGCGACCGCGGTGT 9238
 Qy 387 AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
 Db 9237 CGCGCAGTGGCGTGGCGCAGATG-----TTCAATCGCGCGCGCGGAGGGT 9187
 Qy 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLys 426
 Db 9186 GCCGCTCCGATACCAACTTCGCCAG----- 9160
 Qy 427 ArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle 446
 Db 9159 -----CTGTATGGCAGCGT 9145
 Qy 447 GlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAla 466
 Db 9144 CGCTGTTTCGGG-----GCT 9127
 Qy 467 GlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla 486
 Db 9126 GGC-----GACAA-----GGCCGGTTGATCCTCGCGGGGAGGC 9091
 Qy 487 GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
 Db 9090 GGCACCACTGGACCGACGCTGGTGGCGATGCGCGGACGCTCGGCTTCTCGCGCGT 9031
 Qy 507 GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly-----LeuAlaGlyPro 524
 Db 9030 GCGCCACACAGTTCGGGTATGCTTCGCGGAGTGGCCCGCCAGCCGCCCAAGCCG 8971
 Qy 525 AsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGlyTyrGln--LeuP 544
 Db 8970 GACGAATTCGGCTGGGTGCCAAGAACGTCGTCACCGCGCAGTCGCGGATACGAGCATTC 8911
 Qy 544 roPheThrArgThrLeu--SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsn 563
 Db 8910 CTCAGGCGCGCCCTGGGGCGGGCGGTGTCTCGATGCGCGCAGCGCTTCATGAC 8851
 Qy 564 PheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeu 583
 Db 8850 -----CGGCGGACTGGCACCGGTATCGGCTTCGCGCTCGCGCTCGCGGTG 8797
 Qy 584 AlaProPheSerPheAspIleAlaTyrGly-----HisSerAspLysLysIleArgTrp 601
 Db 8796 GCGCGGTGGCGGTGATATCGCATGCTTGAAGCAGCCCGACTCGCAATTCAGCTC 8737
 Qy 602 HistLeuSerLeuGly 606
 Db 8736 TACATCGATCGGG 8722
 RESULT 32
 LOCUS AE012531 11870 bp DNA linear BCT 23-MAY-2002
 DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 439
 of 460 of the complete genome.
 ACCESSION AE012531
 VERSION AE012531 AE008922
 KEYWORDS AE012531.1 GI:21115353
 SOURCE Xanthomonas campestris pv. campestris str. ATCC 33913.
 ORGANISM Xanthomonas campestris pv. campestris str. ATCC 33913
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xanthomonas.
 REFERENCE 1 (bases 1 to 11870)
 AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,
 Chamberg, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meldanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C. de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
 Kitajima, J.P.
 Comparison of the genomes of two Xanthomonas pathogens with
 differing host specificities
 Nature 417 (6887), 459-463 (2002)
 22022145
 12024217
 2 (bases 1 to 11870)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,
 Chamberg, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meldanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C. de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
 Kitajima, J.P.
 Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
 São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,
 Brazil
 FEATURES
 Location/Qualifiers
 1..11870
 /organism="Xanthomonas campestris pv. campestris str. ATCC
 33913"
 /strain="ATCC 33913"
 /db_xref="ATCC:33913"
 /db_xref="taxon:190485"
 /note="pathovar: campestris"
 480..2099
 /gene="XCC4080"
 480..2099
 /gene="XCC4080"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /translation="polyvinylalcohol dehydrogenase"
 /protein_id="AAW43301.1"
 /db_xref="GI:21115354"
 /translation="MSLDTAFASQIHRSEGSNLPFPASPNNWTFACGGYFNSRFA
 PGWKINRTTAPTAPLTTGDISATVSGGTLVYDPMGGLTLYSDTQSGKPNW
 OAKLSYTGNSAVTNPATISGKSIIVGQASQVYMAIDKTKGLLWSSVAESNPOA
 RITSPVYIGNRIVYTGSSGDMGLAANHTFSRGSVALDNTGKVKWQFYVNPBCY
 TGASVWGQVAIDPISRRLYFGTGNVNPAGVNCVSNRYSGLSTVDELNCNAP
 DNVYDSVALNLDTGLAWADRAAGYDANNLSCIVAPTCNGLCPNTGTPDYDFG
 GAGVQLTVWRDCKPQOLVCGAKGSKIYHAFDARTCAKVMSTOVGCGTGGVWGCTA
 FDPYKGVYVAINNSNTLTGPDNTVSHNGSAAADDAATGATKIRQVAVPQVNSVN
 PNYPPAGCGSLTIVSNLLYAGSMAGNNVALDITGATLWNNYNTATGVSIVSSPVDGAL
 YWGAGYGRNFQTAATSNQLIKFVPSGAAK"
 complement(2282..2587)
 /gene="XCC4081"
 complement(2282..2587)
 /gene="XCC4081"
 /note="putative; ORF located using Glimmer/Genemark"

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM43302.1"
/db_xref="GI:21115355"
/translation="MALARVICLGHWSRWNINAAFRFRAMSTLAPERQLLIHYFADYE
GYCDGRPVLGQLAYSEVEFADGVLDAELVCALRRRAEAHGLEAGQINLCOVTRL"
complement(2490..2930)
/gene="arod"
/feature="arod"
/note="XCC4082"
complement(2490..2930)
/gene="arod"
/note="arod"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="3-dehydroquinate dehydratase"
/protein_id="AAM43303.1"
/db_xref="GI:21115356"
/translation="MILRGPEAASSLVRGPOALPRAVITQLVRCARDAGKTVKALRACG
SEQELDALRVAGARQALIDPGSCVDSVRLHRVLGALAYPYVEAHDDSTDRPERC
LPAGLQQRVATRCYCAQSYLLGLEIALEHLGCTEIQGDVHVCT"
complement(2950..3372)
/gene="XCC4083"
complement(2950..3372)
/gene="XCC4083"
/note="XCC4083"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM43304.1"
/db_xref="GI:21115357"
/translation="MCILIRGPOSDAIPVAPILPCAGRDRLTLACADLASLIAGLQA
AGGSDVELVLLDSGDLPLHRTCAATRAAVDALPTPYIELHADAGQELPWLHPQH
APLAVVITPHDAPRAYAMSLGIARCLPALHLMRVAA"
complement(3616..7572)
/gene="XCC4084"
complement(3616..7572)
/gene="XCC4084"
/note="XCC4084"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM43305.1"
/db_xref="GI:21115358"
/translation="MALAGSASAGGYCARLERPRFAYPALHRHGPVNTTTPPTSPA
PRPRFYRRWVGSGTLVIGGLGALLIWLQTVAGRDVLLAQTKRLPVGASFT
YGSVEPVAGPLTUNVDVRYQDTHFAERYILEPDLRPLGLLRKLQDQAVQVSNLTN
LGKSDPELTPSPESLPQINVPILAQDHIGVDNLRITQLQOPMITLHRVQGLGLEYA
TGEILTRNLVDYDQDFRLNGDYVPDDEYRADLTATAVLPAAAGRTPPASGLVARGD
LDKMEVALAGRAPALRASLVFTCGDDPTWAFNAVTEALDTSLLIPAPAGQETPPATP
IALDLQASGKGGKADLRGSKYQGELSATLOPSPHVSVDQVLTVEPLIIDAFESQTOLR
GSADPRADNARSFRPSVNASGLRTPAADPTDPAVPVELTDARLUGLAGTLKAWAA
IGRATVQRDQGAELVDFSRGNDQRAQLKQVAKTPGGSLLDTGCEVAPALQMDLTA
QLAKFDPGYFQVGNLNGSQVASKRQLPAPAGGVSPGYEATDITPLTGQLRQAL
SANGFALRGFTQGEQOLAGLSRRIIDAKGTVPQDQIDIAAQLOPLQDLDDLPGATGV
RGOLQVNCRRDPATVTDLACTGIRLWMDYSQAQSVLEGRPLWNSGDLQALQGTATGA
GWLDSVRVQARGAEALRLDADTANSWASVALQGVRRNGARQGVQSTLRIVPARG
DAWALRPAQISTGGAFTLSDTCLGAATGALCASANPREGVHGGALPLSLPARG
WLPKQEGRIYTLRDLSDGSGFKPRGNWEGSFRIASPEGGVRLGNRYAAVAGNRP
GELLRYDQVQAQDFTPOQIQGKLGQVFGQNGFDVAKFTGMDAYAPLNGELYLNMSR
LYLLVELVDVVRPKGLVVEGHVSLRGTRDKPLIGGDATLSDTAEYPSMGLTSEGGK
RFDALPDGSAKITASAKSGDGLTNDGGLSWFGTSTPLLLNIRGENVLAYNTSELRIY
ANPDQFGLVITVQLRGKVTVPADIDLERLDRGTSVSEDVVVDLPDPEAPASP
DMDLAIVLGDKNVMSFGGLGGLSGQNLIRARPCEPRANGDLDVRGKYACQODLTI
SRGQITWNNTVSDPFRSLAERKIDGVTAYISGRAESPRADVWSEPMASQSELS
YLVLRGLSTQSSASTSQVSAALSGAGSLIASQIAGKGLDEAGVSESTVGSVGV
GFQGYLSPKLYGVGSMVSGSVLTLKYLILSRGFDVEAESSTVETKGSINWRREK"
complement(7470..9410)
/gene="XCC4085"
complement(7470..9410)
/gene="XCC4085"

```

```

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AM43306.1"
/db_xref="GI:21115359"
/translation="MPHRRRPDLRJCEDLCMYVHVDVHAYTHGMALSTPRRPIINRPS
RSRYPAWRFLILRVPRLVLLILLAGSAFARGTIDKIEIKGIDGDDADMENIEVTS
LSUYEAVGEGESRLEYLLAAOEROTREALPEYGYVSTPIDAAPRDGRLVITRRLAE
NGEYVPQSHSTLSITGMAEODRYLGQDLKQFREGQVFSHPQYEAASKVRIITRLAE
RGYFDADFTQRRVYATRAHADIDLNDWSCRRYDMGTVPFDYDIFRPGLEFLVYVSG
EGSYIHGKLDLRESLSTIDYFSTIDIQPEEADQGRVPVDVILKRAKRTIYVTS
LSYSGAGAGCGVGGVRRYVNSRSGKQITQDYAQNKSLLTTSYRVPARFMDGWTA
SARLYDEQTEYIDLRNVKLTGSRSGQINERWSAIASINALRWRFSGSDFLRMFLGA
TSTLYPQLQANVYVNDRLFPSSYSGOMFIRGGAEGAGSDTNFGOLYGLQALMFLGA
DNGRVLLEGGCTTWSLDAVPSLRFEPAGGSPRIrgvAFREVGPRTPKPDEPFLG
AKNVYTASEYEHYLGKPPGGAVFVDSGSAFDDTDPDWRTGICGFLNWRSPVGPVRVD
IAHGLNDPDSQFOLYIDIGANL"
/complement(9598..11694)
/gene="glys"
/note="XCC4086; glyS(B)"
/complement(9598..11694)
/gene="glys"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="glycyl-tRNA synthetase beta chain"
/protein_id="AM43307.1"
/db_xref="GI:21115360"
/translation="MSGQLPLLIETGTEELPVRKALPGLQAQFDGVLGLEKRGVAVT
RGDKALTPSRRLVALLPGVATPQEPQSEVLGPNLTALDAEGKPTRALAGFAAKAG
IDWTLERSDQAGEBFVRHVPAGQAALLPEILREATAAMPKPMWGAHEYAF
ARPVQMLVLLGGTYVPAELLGVGRDRITGRHFMHGDGDIALAAPGDYIDLRAHVL
VDADARRAIVEEDAAARQAGSARIISDNLBOVNVLMVEPMSAVLCSFSEFAFLAVPQ
EALITENIKNEFVPLDDGGKTEQFIGIANIVDSQDAEVAKGYERYVIRPRFADKFI
FFDDELKQGLEANGAAYTYQAKLTQADKVAARVAALAEAIAPQAGADVPQARRAA
ELAKNDLQSMVNEFPELOGITAGRHVAKAAGEPSETSLAIDEAYQPRFAGDDTALSPL
GKVLATAERLDTLAGGFAAGLKPTGNKDPFALRRNALGLARTVIESGFOLDLPLKIDV

```

Alignment Scores:

Pred. No.:	1..13e-09	Length:	11870
Score:	291.00	Matches:	143
Percent Similarity:	39.16%	Conservative:	90
Best Local Similarity:	24.03%	Mismatches:	270
Query Match:	9.17%	Indels:	92
DB:	1	Gaps:	26

US-09-857-669-2 (1:-609) x AF012531 (1:-11870)

Qy	61	AspSerGluIleLysAspMetValGlucluHsIleuProLeuIleThrGlnGlnGlu	80
Db <td>9137</td> <td>GACCCGAGATGATCGGAACATCGAGGTGCTCTGTGCTGTAT-----</td> <td>9093</td>	9137	GACCCGAGATGATCGGAACATCGAGGTGCTCTGTGCTGTAT-----	9093
Qy <td>81</td> <td>GluValLeuAspLysGluGln-----ThrGlyPheLeuAlaGluAlaPro</td> <td>96</td>	81	GluValLeuAspLysGluGln-----ThrGlyPheLeuAlaGluAlaPro	96
Db <td>9092</td> <td>GAGCCGTCGCGCAGGACGAGGTGAGTCGCGGCTGGAGTACCTGCTCGCGCGCGGAG</td> <td>9033</td>	9092	GAGCCGTCGCGCAGGACGAGGTGAGTCGCGGCTGGAGTACCTGCTCGCGCGCGGAG	9033
Qy <td>97</td> <td>AspAsnValLysThrMetLeuArgSerLysGlyTyPheSerSerLysValSerLeu---</td> <td>115</td>	97	AspAsnValLysThrMetLeuArgSerLysGlyTyPheSerSerLysValSerLeu---	115
Db <td>9032</td> <td>CGGCAGACGCGCAGGCACTGGAGCGGTTTGGTTATTACTCGCCCAACCATCGACATGCC</td> <td>8973</td>	9032	CGGCAGACGCGCAGGCACTGGAGCGGTTTGGTTATTACTCGCCCAACCATCGACATGCC	8973
Qy <td>116</td> <td>ThrGluLysAspGly-----AlaTyThrValHisIleThrProGlyProArgThr</td> <td>132</td>	116	ThrGluLysAspGly-----AlaTyThrValHisIleThrProGlyProArgThr	132
Db <td>8972</td> <td>CGCCGCGCGCAGCGAGCGCCTGACCGTGGTATACACGGTGAACCGTGGCGCGCGGTG</td> <td>8913</td>	8972	CGCCGCGCGCAGCGAGCGCCTGACCGTGGTATACACGGTGAACCGTGGCGCGCGGTG	8913
Qy <td>133</td> <td>LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla</td> <td>152</td>	133	LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla	152
Db <td>8912</td> <td>CGGTGCGCGCAGTCTCATATTTTCGATACCGCGTGGCGCGGAGCAGAC-----</td> <td>8865</td>	8912	CGGTGCGCGCAGTCTCATATTTTCGATACCGCGTGGCGCGGAGCAGAC-----	8865
Qy <td>153</td> <td>GlyTyTyArgAsnAlaLeuGlnProValGlySerAspPheAspGln</td> <td>172</td>	153	GlyTyTyArgAsnAlaLeuGlnProValGlySerAspPheAspGln	172

EOEAQELTRIQERORLNRRTVABEALGELSSTLNSQDQDFPLEGTPLLVAAGAYGRAM
GIKISPLRSNLRNKRVEPLEAIVRASIRMRRLVLLRDNWMLKADGALVADYTKENQPV
ALLPAGNRIETFDVNTYSRVVNEGVAATLAPVAFYFVRSIPDKALRAIDVYKFTLRG
HGMDLQILPTGIFITILLGMITPVATSIIDINAIPDSRSLLOIGLGLVLAAGLAGL
FQLAQFALLRMETSGBASTOAGQVMDRLNLVPVSFFREYTTGDDLRSISSVSAIRRQL
SGRTLIFTSIFALFYQLQFYNYKLALIVASAVVAIAITTSIGILLLSKSVLRQL
EVOCNIFGOTVOLINGISKHIAGAEREAFAAMSKNTYRQIKLELSQLVDFVFIEN
TIIPITSGVLFWFTTKLLEAOTSGSIASFSLGTFLAFNTAFSNFTRGTTNLSNVTQ
ILQIPQWORTOPILTTIPEVNLKSDPRLAGKITVDNVTFCYRRDGNLTLDNVSIS
APEGFEALTGSSGSKSTLLRLLLGPETPOAGSIHYVDGDLGLDVAVRGLGVNL
QNGQLNASIPENTAGDAQITLDEAWEAARMSGAEDVTAMPQMHVTVSEGGNLSG
GQRQLLIRALAKRPRILLFDEATSAIDNRTQAIVSSEIDKLQVTRIATIAHRLSTIR
NAHRIYVLQGRVVQOGTFQELANTEGIFAQUMARQMA"
/gene="all2021"
/complement(3819. .5957)
/gene="all2021"
/complement(3819. .5957)
/note="ORF_ID:all2021"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="BAB73720.1"
/db_xref="GI:17131112"
/translation="MKTPTLLQMEAVECGAALGIILGYGRVVPLPQLRRDCGVSRD
GSKASNMVKAARTYGLKAGFKKELKQVQQLRPPIIVFNFNFHVVVEGFGKHQVYIN
DPATGPRHVSQIEFDEGYTIGVLVMEPGAFOKGRKPSMIASLWSRLOGAESALVYC
IVAGFELTIGLIVPVFSQIFVDIILVQGRMLWHPILLAMAJAALQSGSTLLRLRY
LRLKIRLAVGSSRFLNHLRLVPYSYQRFAGEISNRTTLNDQVADVLAGQLATTI
LNDLFLVYALVQVQDWMVLVMSFAAVNLVLTQWISQRIDANORLQIEYKAAG
ISIAATLQSIETKASGLSEDFSRMSGYTKALNSQOELGTNQTFSVLPITLLSALA
MLLIIVGFRYMDGSLTGMIAFQGLMSGQQPVNVLNFGTTLQLEGLNLRDLDV
LDNVPVEGGTSSSSQALPKIAGYVELKNITFGYSRLAPLTENELSKPGORVA
LVGSGSKSTIAKLVSLGYPTGELYFDSPREELSIQLTNSVAMVEEILLFGG
TVRNLTLWDTTVFDKMLRACQAAIADVIFSMGSGYDAELIEGAMNSGGQRLE
IARALNLTWTVLWDEATSAIDAEKTIIDQNLRRRGCTCIIVAHRLSTIRDCQIIV
LERKVVQOQTHQELWQEGAYSRLRTEG"
/complement(6028. .7395)
/gene="all2022"
/complement(6028. .7395)
/note="ORF_ID:all2022"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73721.1"
/db_xref="GI:17131113"
/translation="MQVSPKSWPLVALGSIIVGVAFTWISYGIPTIVBGRGVLIYP
SNVYPLQSKSAGQIMALNVGVDVIRKQGVLATIDQAEIRKQLOQOQKGLTQLESQDK
AVGSLQRRLEQERSQOQRYLRQRILELEAITPLTKGNTSVDOQROGLQORLQ
QAQALTAVFLRMKTRQELFOKGAISGDEALKAEQEVYONLEKTIADIOAKLKELDVK
ETQOEKDYRENLTIVADLQSQLKOLDQOASVAQOQDLENVTARKKEIOEVKREIGSLR
LQLDNSQIISQHSGRILEITLTPGVQVNAQTRLATIEAENPQSKLVGTYTFPVAEGK
KIQPMGSIQTPQTKRERFGGIVGNVTSISRPITREAAANEVNGNSEVLGVSQQQ
GLIQVFSLEIDANTPSPGKRWSSSTGPHLNTSSGTITTVVRVKVEERAPITFVLPIILRS
ASGIY"
/complement(7472. .9502)
/gene="all2023"
/complement(7472. .9502)
/note="ORF_ID:all2023"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73722.1"
/db_xref="GI:17131114"
/translation="MSDVCLVLMPLYAVERPSIALGGLKARLQAKLEPTIILYINQIF
AEEIGLDVYAMISAFTEFLGEMTFSGVLFPFEPNPNHSEYFDLSPCFKSYENYLPD
RNLKAFIAVREOATFATESVAQSVANLPRIIGCSSTFOQHACASALLKRYRELAPD
IITPMGANCSCAMGLATHTFESWDFIISGEGDLENEFCYQLLDGGREVDPSILPY
GVISPSHRTSKISTASAPASVQNDQVPIDYDDYFOTLTHASTLSPYIQPGLLIETIS
RGCWMGQKHCTFCGLNGSINRKSQDRVLKPTFCRLCYRGLURKQVVDNILDMMH
MNTVPIVLAAGDEAYNIETKANLRWEHLQALQAGVRCIQPGIESMHNSTLKMNK
GNTAMNVLQLLKWARLGIIVFWNLADLPGESDEWYVEMMEWLPILIAHQPPSGASE

IRYDRFSPYVERPSDWGNLNSPLRKYKYVYPLAPEILSELAYFFEDQNNLVRYATEE
NEQTSERTEPGLENLKWVQNMNRLWQSLYKHKPEYLPILSMNDGSGSLTIYDSR
PCATESLLEGLAYWYKCDRALTYREVVSSEINKKLDVPVSNDEIFPIVSELRRK
ILLIDGRLLSLAKGSCSPYLSVSELSPAGYIDLHEFRHNSKSEKVVLLPN"
/complement(9616. .9909)
/gene="asi2024"
/complement(9616. .9909)
/gene="asi2024"
/note="ORF_ID:asi2024"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73723.1"
/db_xref="GI:17131115"
/translation="MSLVMTWKELEEKIIVKAGODESKFQALVSNPRSTLEKEGINLP
SSIDVNLVETTPGNLSQALPNSEQDTLSEALESVAGGRISGEVSVTVKVF"
/complement(10195. .13482)
/gene="all2025"
/complement(10195. .13482)
/note="ORF_ID:all2025"
similar to Mrsd protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73724.1"
/db_xref="GI:17131116"
/translation="MNVSKQSLIKIIRSSNISBELSSSFIENDQINANONNAKLEK
WYQFIAGQNHNEFKRLWMDGLNTISCVIGNVNLADEKYLPMNVEFLKAGMQVATF
VIPSNNKFLQOHRYLDPERPFEEILVPFVDAKKKLTESAGSSYVLLISQNAHIQ
MERSLHELTGICISPLEFISFRLSSPLVPRIQOLNSCKSGYQTFVFNKLLKD
GLLSFQBYSVLARLIATVDFVWEATGEFSLRLASDSDKIQLTFQPOKLGQVVSVO
LAMDENRGRRAVIVIPASGFKLVYKPKSLGLEKAYFDFLDWLNQOKINLPKLKI
INCNSYCHMEAEALPCQDQADAVKHQYQACWVLCIYVLLAGNDCHCNLACGDQPV
LYDLETLLHHTTWLSKDDADASTANECIQDSVVGTFPGWQILPYEQTELKILDFS
GLGFGQEGMEYRANKWKHINTDSNVIVREYTKLLPKRPFGENIDTSLNNHSELI
DGFQMTQDLOQKREELASDPITAFNSQKRVLRINTVYATSLQNSLHKCLRTG
VEHSTIQDLSRAFSLSEDDKHPMLPAAERKALQDLPYFTAYSDSNVIDISEQT
INKFLSSSYDDVTAHLRQMDADLAKOISLIRGSLYSCLTNEOLNSUPLENPNLYMD
IVAPTERGMLOQATAIAQOLEQOAIKRGKNSVTWIGMYLFEAGRFOLQPVGNLSYD
LASIYVALRISFELGEVELIQDAQLASLWVINEATVQAPVMTVAGGTILSLAL
YKATKEPATLQQAISWGNLYLDTRVYDEVYQAMTNAEGKVLGFSHGAAGIAYALLQ
LYATLDEPIFASAREAITLEQRFLSSNCQDLIDSIDNSLSSSTFEMCSGSLIGIS
VARLGTNLIDOTEIREOLDNLHTIHRSSQLVDNICCGNFSEHTEMLVASKOLSHR
ELLETTANNRATQVLSRAQIGYFQLLPHLPADVYSFGFCQGIAGIYELLRLSCFDLL
PSILLWQ"
/complement(13877. .14227)
/gene="all2026"
/complement(13877. .14227)
/note="ORF_ID:all2026"
similar to nitrate hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73725.1"
/db_xref="GI:17131117"
/translation="MSEQTKRKEIEAQIIVQAWKDETYRQELLNNSKAVIEREFAIQ
LPKEINHVVEENDSNFYFVIPARPNDLSELSQLEAVAGGSLGDIFTFVEKTYDR
LKDFGEIYKAIVD"
/complement(14385. .14720)
/gene="all2027"
/complement(14385. .14720)
/note="ORF_ID:all2027"
similar to nitrate hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73726.1"
/db_xref="GI:17131118"
/translation="MSEQTKRQVQEAQIIVQAWKDEAYRQELLNPKKIVEQEPGVG
LPEGTIVHVMENASNLFYVIPARPNDLELSDQLEAVAGGALLWTLLLLPIAHG
ALEEHSNRK"
/complement(14854. .15042)


```

QY 586 eSerPheAspIleAlaTyrGly---HisSerAspLysLysIleArgTriPhisIleSerLe 605
||||| :|: ||| ||| :|: |||
Db 320739 TCCTATTGAC-----TATGGTATCAACGATGACGGTGATGCCGATCAATTTCGGTAT 320792

QY 605 uGlyThrArgPhe 609
:|: ||| ||| |||
Db 320793 TCGAGAAAGGTTT 320805

RESULT 34
AE005835
LOCUS
DEFINITION
  AE005835 10301 bp DNA linear BCT 12-JUN-2002
  Caulobacter crescentus CB15 section 161 of 359 of the complete
  genome.
ACCESSION
  AE005835 AE005673
VERSION
  AE005835.1 GI:13422995
KEYWORDS
SOURCE
  ORGANISM
    Caulobacter crescentus CB15.
    Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
    Caulobacter.
  1 (bases 1 to 10301)
    Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
    Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
    Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
    Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
    Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
    Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
    Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
    Fraser,C.M.
  Complete genome sequence of Caulobacter crescentus
  Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  21173698
  11259647
REFERENCE
  2 (bases 1 to 10301)
    Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
    Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
    Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
    Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
    Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
    Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
    Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
    Fraser,C.M.
  Direct Submission
  Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
  Medical Center Dr, Rockville, MD 20850, USA
FEATURES
  source
    1. .10301
      /organism="Caulobacter crescentus CB15"
      /strain="CB15"
      /db_xref="taxon:190650"
    83..439
      /gene="CC1599"
    83..439
      /gene="CC1599"
      /note="identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAK23578.1"
      /db_xref="GI:13422995"
      /translation="MVMVETSGARRHPRMPAARKIYIVDDPRSKASLLDVAEKGK
      RISIAGIASPPDTEFVDAGRRVHLANVMVRSCTEVGVQFAATQIGRPGAGAGAL
      ETARRPLATLPADDDA"
      complement(538..1515)
      /gene="CC1600"
      complement(538..1515)
      /gene="CC1600"
      /note="identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="AAK23579.1"

```

```

/db_xref="GI:13422997"
/translation="MLLEIRHVTVQHYERPVPVSESLMELMWMQKTKTARORLVSGFELDNLN
PAOVFSEADSEFGNAVYHEDVPOPHNLTLIARSAVETESPCERDQDQDMGEMDLRLS
EVVRECEDFLRPHGFVTRTEALQTFINEHDDLDALKRRDPLTAVSLSTIYIGAFEYQ
PGVTDADSPIDLTLSAGRCVQDFAHIMILAVCKGVMGVPARYVSGYLFITDRDAGDSQD
DAYHAMVEVFLSPRLNMHGVDPNTNMHAGVHVAVAGRDYADVTSRGVYKGDASEQL
AVGVSRRARAALAEPEFLMARPSFAGRRRRHTALREELHQOQOQOQO"
complement(1527..2474)
/gene="CC1601"
complement(1527..2474)
/gene="CC1601"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23580.1"
/db_xref="GI:13422998"
/translation="MMIARVADSLYWLGRYIERAEHLKSLSVMLNATLDOTDAGAAQ
WITAMAAVAPEDGAGVGSFEAAHALVLDSDPNVSVSSLARARENARQVRDQITET
WERLNLVLYKVTDPNATDFAAGSDIFLHDVIADLHLFKGAADTMSHGSESWRFMMLG
IYMERAGLVSRLLVECVFAESPTHGRIGHDHVALVSVLRMACALEPYLVRVYTAETEPHHI
LEFLVFDEDFPRSIRFATAQIEQHILSAMARSVSANERACPERLAGRLKARLOFADVDE
LEAVGAGPLTTVNECARCIHEAIVETVAYPLETRLPA"
complement(2471..4108)
/gene="CC1602"
complement(2471..4108)
/gene="CC1602"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23581.1"
/db_xref="GI:13422999"
/translation="MRNLADVESGKQERVEALESGRRMPLGQAQRDRROADRSVAAKT
QTLADTPTLPMTAAALPGVAYDEMFTPEGEVRQHYDPLHGRMSTLGAEELAAQRRTL
ERSFLQGITFTVYGAENTERIITDLEPRILIPASEWARIAGLQRLRALNLFLLAD
IYCDQOILMDGVVPRELVGAPSRREMOLHLYPHKAYANVGSDLRQCDQGFAYLE
DNLVPGVSNAYEHAYLARLMGVPLVEGRDLVHNVMYMTTGLRIDVIVRRY
DDDFIDPLTFRDSSSLGAAGLNNAYRAGNVICNAPGTGVADDKAYAYVPDIIRYL
GEDALPNIEITYLCREPRQLSHVLNLDKLVKAVGSGGYGLVGPASQKQERDDFA
LAIKADPENYIAOPTIQLSTAFCLVDGRIEPRHVDLRPILSGEKTIVTPGALTRVAL
KRGSLVNSSSGGGSKDTWVLADDAAPSNGHGSARP"
4142..6028
/gene="CC1603"
4142..6028
/gene="CC1603"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23582.1"
/db_xref="GI:13423000"
/translation="MRSARDGGGRHSRSDTSPARDYSRGGGASRGERDLVLGRTTIAL
AAVTLWSAANAWEDEMAQIQGVEDRALDRAIQRLSDSKQPPRSRSEARRRQAGE
DAIVLAEGYVAYTVEDVTEGDPRAIVRTTPSPAFLLADPHIDWSGSPDDEGVYRQ
RAVAMRLTEGEPBSADVVGAEGRIYAQVAKIGYADVAEPREVVDHADRTVPTFF
RIMAGELVRLNGVDVYVTKGTNPENWVGLAPWVAGVYDPEOVAELERLRLDRTAVYDS
ISVSLAGTDKSAEAGYRPVVTLSDRRARTIELGAGYSTSEGAGYDARVIRYNQRKA
DTTYYALRFKLEQRLGAEISLPHMRPQOITLKNSSVFRNDTDAYNETGATVGVDLT
RRORTTAYTRTFGVSDLSQTKQOVNRNGLIAGRLKLNLATLAGLAAYAWFSDIILDPK
RWMLRTAEPTVYVAGDTSVPYLLKAGOGSAYLPFGKQDSTVLAARVKKGLAILGAGLL
DVPASRRFSGGGSVRGYAYQAIQPRISDNTPOGGSILVETSFVROKTRDWSGVA
FVADAGIHTHTPQREDFRAGVYRDLGFGPIRADIAAPLGRKKGDPKFOIILSI
GOSF"
6025..10221
/gene="CC1604"
6025..10221
/gene="CC1604"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

```
/protein_id="AAK23583.1"
/db_xref="GI:13423001"
/translation="MSRDQKPDLTDTVSOAAEAAGETAVKVVKRIGWGVVALVYVGV
LAALLMAGGLRLAPITQGRMFLEARASGLKLRIGLKHIEGLSDTWDRGVRRL
TISDEKILEARELVANRATLFGRRPHAQITARDVTVLRPMTPKPGKSSAPV
SVLDTFKRLTRPAPSAEGDFDVGATEMARRGGQKRIATASRLHGHDLVDVDF
DGRSKSLRVADATBANGATAGLGPADRTFDLKAKEASTASRAAFDVAARSRE
IPLKARGATEOGSSARGRIDLEASRLTQRLARMFPGADFSIEGRKATADGLSALKI
SLNSENLVGGGLADLKKRRTPGDIAFTAOVAGLSRIVTPKMRGAQOAGFRID
GGFVBERMDVSDLELLVLRKATGCPAKVSMKGNLAIEGAATTAGGGGTGLLAVLG
SSPKATFEGRVYKGRFLITAAARIEPNLIVTKGGRGLLGLSFGENKMYASLAPVR
LTKAGSKASSPAGRPWSETVDGRAANFGTYAEVDRLRLGLTPRLAGKANWAG
LVTVRLARLDGANAALTKSKLGQGLALDFGVDMWTATGPPRAGPVPVAGQAGDGRVG
GTLSPALQADFTIELPRLPLSKARLTLPARGPNTGDAAALQADSOHQPARGR
ASPRFPGGTIELTVDVDAGGAKAQSALTRDGPATADLOVAGPACFLPGCTVAGS
ARLVADLGGAPSATLDLVAGVATGSMKVRSAIKADGPLSLPLSIDARGPAGGL
WRLSGADLAQTKIKIYDLTLNASGRVTRTELKTRDAARVFGDGPQOARLRLADGR
ADIDAGGATAMKLAELAGLITLAFDPLDGDIDAFNLQGGDGLVPEASAMPRLAVN
RERKATIESLDKTIQGLTSDSOLTIIADGNNRQGLKAEANLVPVEASAMPRLAVN
RRRGVQGRSAQGEIKPLMDLLGPERSLGRINLQASLGTLADPLRIGQATLDGGG
FDQCSGLRLRNVTLSALADNAIDVSOAIGEDGGSSISGAGRISSLRGGEGNFELD
MKSFRILDNLASAVATGOASVNRSDGKIKLTGAVTLDRADVSAOTKYTPAGVAMDV
VERNRPVLDLQGLQRPTTGGMILLDLKAPRIFVYRGRGLDVELSDAHVGTSLAP
QLSGVARMVGETDFAGKRFEPDDGCVYLSLQDQVRDLDSATREDTALTAVVRIG
TAAKPEITUTSKPELPSDEVLSQVLFSGASAAQLSPTEAAQLASALAAAGGGGFDVIG
NLRSFARLDRLAPAEATGWTAGCKRYVTDVYVEIIGGRDGPAAQVWEVRIIRLTSL
VSRIGGNDAKLSVRWKDY"
```

BASE COUNT 1623 a 3259 c 3692 g 1727 t

ORIGIN

Alignment Scores:

Pred. No.:	1.65e-08	Length:	10301
Score:	272.50	Matches:	138
Percent Similarity:	38.59%	Conservative:	92
Best Local Similarity:	23.15%	Mismatches:	285
Query Match:	8.59%	Indels:	81
DB:	1	Gaps:	19

US-09-857-669-2 (1-609) x AE05835 (1-10301)

Qy	60	GlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln	79
Db	4343	GAGGACCGAGCCCTTCGCGACGCGCATCCAGCGCCGCTG-----TCGGATTCCAAA	4393
Qy	80	GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnVal	99
Db	4394	CAGCGCGCGCAGCCGCTCCGAGCGCGCGCGCGCGCTCAGCGCGGTGAGGATCGG	4453
Qy	100	LysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal-----SerLeuThrGlu	117
Db	4454	ATCGCCGCTCTGCGCGGAGGGCTACTACGCTTACACCGCTCAGCGCGGATGTGACCGAA	4513
Qy	118	LysAspGlyAlaTyrThr---ValHisIleThrProGlyProArgThrLysIleAlaAsn	136
Db	4514	GGCGATCCGCGCGCGCATCGTGAGGATCAGCGCGGAGCGGCTTCCTGCTGCGCGCAT	4573
Qy	137	ValGlyValAlaIleLeuLysAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArg	156
Db	4574	CCCCACATCGACTGCTGGGCTCCCGCGCGATGAAGCGTTCGT-----	4618
Qy	157	AsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGlu	176
Db	4619	-----CAGCGCGCGGTTGCGCGCATGCGGCTTAACCGAAGCGGAGCGCT	4660
Qy	177	AsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeu	196
Db	4661	GGGCGCTCCCGCGCGCGCGGCGGCGGCGGATCGTCCGCCAGGTGGCCCAAGCTC	4720
Qy	197	GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----	211
Db	4721	GGCTACCGCGAGTTCGCGCGCGCGCGCGGCAAGTTGTTGTCGACCATGCTGACCGCACG	4780
Qy	212	-----LeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIle	229

Db	4781	GTTCGGCGCAGCTTCAGGATCATGCGCGGAGCTGCTGCTGAAAGCGGCTTGATGTC	4840
Qy	230	ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGly	249
Db	4841	GTCCAAAGGGGCGCACCAATCCGGAATGGTGGCGCTCTCGGCGCTCGGGTGGCGGT	4900
Qy	250	ThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHis	269
Db	4901	GACGCTCATGATCCGGAAGATGTGCGCGAGCTGGAGCGCGCACTGCGCGACACCGCGT	4960
Qy	270	TyrSerGlyAlaSerValGln-----AlaAspPheAspArgLeuGlnGlyAspArg	286
Db	4961	TATGACTCGATCTCGGTTTCGCTCGCGGAGCACCAAGGCTTCGCGCGAGGCTATCGG	5020
Qy	287	ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg	306
Db	5021	---CCGCTGTCGTCACCTTGTCCGACAGCGCGCGCGCATCGAACTGGGCGCGCGG	5077
Qy	307	LeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn	326
Db	5078	TATTCGACGACGAGCGCGCGCTCGATCGCGCTGGATCCGCTATACCGCCACAAG	5137
Qy	327	LysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAla	346
Db	5138	CGCGCCGACACGACGACATATGCGCTCGCTCGCAAGCTGGAACAACAGCTTCGGCGCA	5197
Qy	347	GlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer-----	361
Db	5198	GAATCTCACTCGCG-----CACTGGCGGCGACCGCGACAGACCTTG	5239
Qy	362	-----AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer	379
Db	5240	AAGCTGAACAGCTCGGCTTCCTCCGCAACGACACCGACCGCTTACACGAAACCGGCGC	5299
Qy	380	GlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArg-----LeuGlyAlaGlu	398
Db	5300	GTCGGCGTCGATCTGACCCCGCGACGCGCGCGCTATCGACCTTCGCGGTTCCTCA	5359
Qy	399	Phe-----LeuAlaGluGlyArgLys	405
Db	5360	TTGACACTCTCGCAGACCAAGGAGCGGTCAATGCGAACCGGCTTGATCGCGCGCGGAAG	5419
Qy	406	IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu---ThrAlaSer	424
Db	5420	CTG-----AATCTCGCGACCTCGCGCGGTCTGCGCGCGC	5452
Qy	425	TrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGly	444
Db	5453	TACGCTGGGATTTCTCGGACGATATCTCGATCCCAACGGGGGTGCGCGCTGGAGACC	5512
Qy	445	LysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla	464
Db	5513	CGGCGCGCAACGACTTATGTGCGCGGCGACACTCTCGCTCCCGTACTTGAAGTGGCTGGC	5572
Qy	465	ArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIle---IleArg	483
Db	5573	CAA---GGCTCGGCGCTATCTGCGCTTTTGGCAAGCAGGACGACCGGCTGGCGCGCGC	5629
Qy	484	GlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPhe	503
Db	5630	GTCAAAGTGGCGCGCATCTTTGGCGCGCGGCTTCTCGATGCTCGCGGCTCCCGCGCTTC	5689
Qy	504	ArgSerGlyAlaLysSerValArgGlyTyrGluLeuAspSerIleIleGlyLeuAla---	522
Db	5690	TTCTCCCGCGCGCGGCTTGGTGGCGAGGCTACGCTATCAGCGCATCGCGCGCGCTTG	5749
Qy	523	-----GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu	539
Db	5750	TCGACAAACGCGCAAGGGGGTATT-----TCACCTGGTCGAGACCTCGCTTT	5797
Qy	540	GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp	559

```

Db 5798 GAGTGGCCGAGAGATCACCGATCGCTGAGCGCGCTCGCCCTTCGATGCGGGCGCT 5857
QY 560 AlaAlaAlaAsnPhe-----LysArgMetLysLeuLysHisGlySerGlyLeuGlyVal 577
Db 5958 ATCGGGACCATGAAACCCCGAGCGGAGGACTTCGCGCTGGCGCGCTTGGGCTC 5917
QY 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAla-----TyrGly 593
Db 5918 CGTTACGATCTGGGCTTTGGACCGCATCCGGCGCGGACATCCGCGCGCTGGGAGCGCG 5977
QY 594 HisSerAspLysLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 5978 AAGGGCGACCGGAGTTCAGATCTATCTCAGCATCGGCAAGACTTT 6025

RESULT 35
NGU81959
LOCUS NGU81959
DEFINITION Neisseria gonorrhoeae outer membrane protein (omp85) gene, complete
cvs.
ACCESSION U81959
VERSION 1
KEYWORDS
SOURCE
ORGANISM Neisseria gonorrhoeae.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 2379)
Manning,D.S., Reschke,D.K. and Judd,R.C.
Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis
are similar to Haemophilus influenzae D-15-Ag and Pasteurella
multocida Oma87
Microb. Pathog. 25 (1), 11-21 (1998)
98379445
PUBMED 9705245
REFERENCE
2 (bases 1 to 2379)
Reschke,D.K., Manning,D.S. and Judd,R.C.
Direct Submission
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
/organism="Neisseria gonorrhoeae"
/strain="FA19"
/db_xref="taxon:485"
1. .2379
/gene="omp85"
1. .2379
/gene="omp85"
/note="omp85"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AAC17600.1"
/db_xref="GI:1766042"
/translation="MKLQIATASALMMLGSLPFADEFIQQDIRVEGLQTEPSTVFN
LPKVGDTNDTHGSAIISLVAITGFDVVRVETADGLLLTVIVGPTIGSLNITGAK
MLQNDIAIKNLSFGLAQSLFNOATLNOAVAGLKEEYLGRKLNQINPKVTKLARN
RDVIDITDGGKSAKITDIEFCNOVYSDPKLROMSLTEGGIWTWTRSDRDFRQKF
AQMKVETDEYQNNGYDFRILDTDLQTHEDKTRQIKITVHEGGRKGVKSIEGT
NEVPAELKLLMKPKKWTROQMTAVLGEIONRMGSAGYSEISVQPLPNAGTKT
VDPLVHIEPKRTIYNEIHTGNKKTREDEVRRELROMESAPYDTSKLQSKREVELL
GYFDNQFADRLAQTGTPDKVDLNMNLSLTERSTGSLDSAGWQDTGLVMSAGVSQDNLF
GTGKSAALRASRKTTLNGLSFTDPYFTADGVSIGYDIYGAQFDRKASTSVKQYKT
TTAGGVRMGIPVTEYDRVNFGLAEHLTVNTYNKAPRYADFIKOYKTDGADGSEK
GLLYKTVGGRNKTDSALNPIRGYLTGVNAELGSKLOYISATHNQWTFPLSKT
FTLMLGGEVGIAGGRTKEIPFFENYGGGLGSVRGESGTGLPKVYDEYBKSIGY
GNKNVSLRMPFCAKDARTVRLSLFADAGSVWDGRTYTAENGNNKSVYSENAH
KSTFTNELRYSAGGAVTWLSPLQPMKFIVAYPLKKKPEDEIQRFQFOLGTFP."
BASE COUNT 636 a 752 c 592 g 399 t
ORIGIN
Alignment Scores: 1.55e-08 Length: 2379
Pred. No.:

```

```

Score: 260.50 Matches: 145
Percent Similarity: 36.55% Conservatives: 86
Best Local Similarity: 22.94% Mismatches: 245
Query Match: 8.21% Indels: 157
DB: 1 Gaps: 28

US-09-857-669-2 (1-609) x NGU81959 (1-2379)
QY 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys-----ThrMet 102
Db 626 GAACCGACCGTTTCGACCGCCAGAAATTCGCCCAAGACATCGGCAAAAGTAAACGACTTCT 685
QY 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 686 ACCAGAACAAAC-GGCTACTTCGATTCCTCGATCTCCGATACCGACATCCAAACCAACGAA 744
QY 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAAACCGAGCGACACCATCAAAATCACCCTCCAGAACGCGGACGCTTCCGCTGGGC 804
QY 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAGTGTGCAAT-----GAAGCGACACCAACGAAAGTCCCC 840
QY 156 ArgAsnAlaLeuGlu-----AsnTrpGlnGlnProValGly 167
Db 841 AAGCCGCAACTGGAAACAACTGCTGACCATGAAGCCCGCAAAATGATCGAA----- 891
QY 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCGCTTTGGGTGAGATTGAG 927
QY 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 928 AACCGATGGCTCGGCGAGCTACGCATACAGCGAAATC---ACGCTACGCCCTGCGCG 984
QY 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAla 223
Db 985 AAGCGCGGAACCAAAACCGTCGATTCTCTCGACATCGAACCGCGGACAGAAATCTAC 1044
QY 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1045 GTCAACGAAATCCACATCACCGCAACAAACACCGCGACGAAAGTCTGCGCGCGGAA 1104
QY 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAla 263
Db 1105 TTGCGCCAAATGGAATCCGCGCTTACGACACCTCCAAAGCTGCAACGCTCCAAAGAGCGC 1164
QY 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1165 GTCGAGCTTTGGGCTACTTCGAC-----AACGTACAGTTTGATGCGCTCCGCTTCCC 1218
QY 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACGCGCCGACAAAGTCGATTTGAACATGAGCTTCCGCAACGCTTCCACCGCTCGCTC 1278
QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
Db 1279 GACTTGACCGCGGCTGGTTTCAGATACCGGCTTGTCATGCTCCGCGGGGCTATCCGAG 1338
QY 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGlu 341
Db 1339 GACAACTGTTTCGCTACGGGC-----AAGTCGCGCGCCCTCGCGCTCCGCAAGCAAA 1392
QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 1393 ACCACGCTCAACGCTCGCTGCTCG-----TTTACCGACCGCTACTTCCACGCA 1440
QY 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1441 GACGGGTGACGCTGGGCTACGATATTTACGGAAAGCGCTTCGACCGCGCAAGCATCG 1500
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381

```

Db 1501 ACCACGCTCAACAATATAAAACACCACCGCGCGCGCGTAAAGATGGGTATCCCC 1560
Qy 382 ---ileTptTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGAATACGACCGCGCTCAATTCGGCTGGCGGCGGACACCTACCGTCAACACC 1620
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 1621 TACAACAAAGCACCACCAACCGTATCGGACCTTTATCAACAACATAGCGCAACACCGACGCG 1680
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu----- 425
Db 1681 GCAGAC-----GGCAGCTTCAAGCGCTGTATCAAAAGGCACTGTCGGCTGG 1728
Qy 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GGGCGCAACAGACGACCGCTTATGGCGGCGCGCGC---TACCTACCGCGC--- 1782
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAAATGCGGAATCGCCCTCGCGCGCAGCAAA 1815
Qy 466 AlaGlyTyrPhePheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAATACTACTCGCGCACCAACCAACACCTGGTCTTCCCTTAAGCAAAACCTTC 1875
Qy 479 ThrPheIleLeuArgGlyAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACGTGATGCTCGCGGCGGAGTCGGCATTCGGGCGCGCTACGGCAGAACCAAAATC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCTTCTTTGAAACTTCTACGGCGCGCGCTGGGTTCGGTTCGGGCTACGAAAGCGCGC 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACGTCCGCGCGCAAGTGTATGACGAATACGCGCGCAAAATACAGTACGGCGGCAAC--- 2052
Qy 527 SerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCCAACTCTCCCGCGAGCTCTTCCCGATGCCCGGTGCG 2100
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACACGCGTCCGCTCAGCCTGTTCGCGACGACGAGCGTGTGGGACGCGC 2160
Qy 560 -----AlaAlaAsnPhelLysArgMet----- 567
Db 2161 AGAACCTTATACCGCGCGCGCAAAACGGTAAACAAATACGGTTTACTCGGAAACCGCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSer 581
Db 2221 AATCCACCTTTACCAACGAATTCGCTATTCGCGCGCGCGGCTTACCTGGCTCTCG 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CTTTGGCGCGCATGAATATCATCTACGCTTACCGCTGGAAGAAAAACCGGAAGACGAA 2340
Qy 598 LysIleArgTyrPheHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCCAACGCTTCCAATTCACGCTCGGACGACGCTTC 2376
RESULT 36
AX155443
LOCUS AX155443 2379 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 6 from Patent WO0138350.
ACCESSION AX155443
VERSION AX155443.1 GI:14536771
KEYWORDS
SOURCE Neisseria meningitidis
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;

Neisseria.
1 (bases 1 to 2379)
AUTHORS Giuliani M.M., Pizza M., Rappuoli, R. and Holst, J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 6 31-MAY-2001;
Chiron Spa (IT) ; Statens Institut for Folkehelsete (NO)
FEATURES
Location/Qualifiers
1..2379
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 636 a 755 c 596 g 392 t
ORIGIN
Alignment Scores:
Pred. No.: 2,87e-08 Length: 2379
Score: 256.50 Matches: 144
Percent Similarity: 36.39% Conservative: 86
Best Local Similarity: 22.78% Mismatches: 246
Query Match: 8.08% Indels: 157
DB: Gaps: 28
US-09-857-669-2 (1-609) x AX155443 (1-2379)
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys-----ThrMet 102
Db 636 GAAGCGACCGGTTCGACCGCCAGAAATTCGCCCAAGACATGCAAAAGTAAACCGACTTCT 685
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 686 ACCAGAACAC--GGCTATCTCGATTCGATACCGACATCCAAACCAACCAACAA 744
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAACACCGACACCATCAAAATCACCGTCCACGAAGCGGACGTTTCGGTGGGGC 804
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAAGTGTGCAAT-----GAAGCGCACACCAACCAAGTCCCC 840
Qy 156 ArgAsnAlaLeuGlu-----AsnTyrGlnGlnProValGly 167
Db 841 AAGCGCGCACTGGAAACAACTGCTGACCATGAGCGCGCAATGGTAGCAA----- 891
Qy 168 SerAspPheAspGlnAspSerTyrGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCGCTTTTGGGTGAGATTTCAG 927
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 928 AACCGCATGGCTCGGCGACGCTACGCATACAGCAAAATC---AGCGTACAGCGCGTGGC 984
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223
Db 985 AACCGCGGAACCAAAACCGTCTGATTCCTCTGCATCAGAACCGCGCGGAAATATAC 1044
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1045 GTCACGAATCCACATCACCAGCAACAAACACCCCGCAGAGTCTGCGCGCGCAA 1104
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAla 263
Db 1105 TTGCGCCAAATGGAATCGCGCGCTTACGACACCTCAAGCTGCAACGCTCCAAAGAGCGC 1164
Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerGlyAlaAspPheAspArgLeuGln 283
Db 1165 GTCGAGCTTTGGGCTACTTCGAC-----AACGTACAGTTTGATGCGCGTCCCGCTGCC 1218
Qy 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACGCGCGACAAAGTCGATTGAAACATGACCTGACCGCAACGCTCCACCGGCTCGCTC 1278
Qy 302 GluThrGlyIleArgLeuAspSerGlyTyrGlyLysIleAlaTyrAspTyr 321
::: |||||

Db 1279 GACTTGAGCGCGGCTGGGTTCAGGATACCGGCTGGTTCATGTCGCCGCGGTATCCGAG 1338
QY 322 TyAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGlu 341
Db 1339 GACAACCTGTTCGGTACGGG-----AAGTCGGCGCCCTCGCGCTCGGAAGCAAA 1392
QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTyrThrSer 361
Db 1393 ACCACGCTCAACGCGTCCGTGTG-----TTTACCAGCAGCGTACTTACCGGCA 1440
QY 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1441 GACGGGTGACGCTGGCTGAGTATTTACGGAAAGCCCTTCGACCGCGCAAGCATCG 1500
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
Db 1501 ACCAGCGTCAACAATATAAACCACCACCGCGCGCGGCGTAAAGATGGGTATCCCC 1560
QY 382 ---IleTyrTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGAATACGACCGCTCAATTTCCGGCTGCGCGGGAACACCTGACCGTCAACACC 1620
QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgGlyIleProGly 408
Db 1621 TACAACAAGCACCAACACCTATGCGGACTTTATCAGGAATATACGGCAAAACCGACGGC 1680
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTyr 425
Db 1681 GCAGAC-----GCACGCTCAAGGCTGCTGTACAAAGCCACCGTCCGGCTGG 1728
QY 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GGGCGCAACAAGACGACGCGTATGCGCGGCGGCGG---TACCTGACGCG--- 1782
QY 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuLeuArgThrSerAlaArg 465
Db 1783 -----GTAATGCCGAATTCGCCCTGCGCGGCGGCAAA 1815
QY 466 AlaGlyTyrPheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCATATCTACTCCGCGCCACCAACCTGTTCTCCCTTAACCAAAACCTTC 1875
QY 479 ThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACGTGTATGTCGCGCGCGAAGTCGGCATTTGCGGCGGCTACGCGCAGAACCAAGAAATC 1935
QY 498 ProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCCTTCTTGAACACTTCTACGCGCGGCGCTGGGTTTCCGTCGCGGCTACGAAGCGGC 1995
QY 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACGCTCGCGCGAAGTGTATGACGATACGCGGCAAAATACGCTACGCGGCGAAC--- 2052
QY 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCAACGCTCTCCGCGAGCTCTTCCCGATGCGCGGTGCG 2100
QY 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACGACCGCTCCGCTGAGCTGTTTCCGACGACGCGGCGTGTGGGACGGC 2160
QY 560 -----AlaAlaLeuAsnPheLysArgMet----- 567
Db 2161 AGAACCTATACCGCGCGGAAACGGTAAACAACAAATTCGGTTTACTCGGAAACCGCGCAT 2220
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSer 581
Db 2221 AATCCACCTTTACCAACGAATTCGCTATTTCCCGCGCGCGGTTTACCTGGCTCTCG 2280
QY 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CTTTGGTCCGATGAATTCAGCTACGCTACCGCTACCGCTGAAGAAAAAACCGGGAACGAA 2340

QY 598 LysIleArgTyrPheHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCCACGCTTCCAAATCCAGCTCGCGACGACGCTTC 2376

RESULT 37

AX202493 2379 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 6 from Patent WO0152885.
DEFINITION
ACCESSION AX202493
VERSION AX202493.1 GI:15392216

KEYWORDS

SOURCE
ORGANISM
Neisseria gonorrhoeae.
Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.

REFERENCE

1 (bases 1 to 2379)
Pizzi, M.C., Rappuoli, R.C. and Giuliani, M.C.
Outer membrane vesicle (omv) vaccine comprising N. meningitidis
serogroup b outer membrane proteins
Patent: WO 0152885-A 6 26-JUL-2001;
Chiron Spa (IT)

FEATURES

Location/Qualifiers
Source
1..2379
/organism="Neisseria gonorrhoeae"
/db_xref="taxon:485"

BASE COUNT 636 a 755 c 596 g 392 t

ORIGIN

Alignment Scores:

Pred. No.: 2,87e-08 Length: 2379
Score: 256.50 Matches: 144
Percent Similarity: 36.39% Conservative: 86
Best Local Similarity: 22.78% Mismatches: 246
Query Match: 8.08% Indels: 157
DB: 6 Gaps: 28

US-09-857-669-2 (1-609) x AX202493 (1-2379)

QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAsnValLys-----ThrMet 102
Db 626 GAACGACCGGTCGACCGCGCAAAATTCGCCCAAGACATGGAAAAAGTACCGACCTCT 685
QY 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 586 ACCAGAACAC-GGCTACTTCGATTCCTTCATCCGACATCCAAACCAACGAA 744
QY 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAAAACCGGACGACCATCAAAATCACCGTCCACGAAGCGGACGCTTCCGCTGGGC 804
QY 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAAGTGTGCGATT-----GAAAGCGACACCAAGAAAGTCCCC 840
QY 156 ArgAsnAlaLeuGlu-----AsnTyrGlnGlnProValGly 167
Db 841 AAGGCCCAACTGGAAAAACTGCTGACCATGAGCCCGCAATGCTACGAA----- 891
QY 168 SerAspPheAspGlnAspSerTyrGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCGTTCGCTGGGTGAGATTTCAG 927
QY 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
Db 928 AACCGCATGGCTCGGACGCGCTACGATACAGCAATC-----AGCGTACAGCCGCTCCG 984
QY 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223
Db 985 AACCCCGGAACCAAAACCGTCGATTTCGCTGCACATCGAACCGCGCGGAAATCTAC 1044
QY 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243


```
QY 126 lethr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
|||||
Db 770 TCACCGTCCACGAGGCGGAGCTTCGTGGGCGCAAGTCTCCATC----- 816

QY 144 spleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
|||||
Db 817 -----GAAGCGCAGCACCAAGAGTCCCCCAAGCCGAACCTGMAAACTGCTGA 865

QY 161 -----AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpG 176
|||||
Db 866 CCATGAAGCCCGCAATGGTACGAA-----C 892

QY 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
|||||
Db 893 GCCAGCAGATGACCGCGTTTGGTGAGATTCCAGAACCGCATGGGCTCGCAGGCTACG 952

QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
|||||
Db 953 CATACAGCGAAATC---AGCGTACAGCGCGTGCACAAACCGCGCAACCAACCGTCGATT 1009

QY 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrClyT 232
|||||
Db 1010 TCGTCTGTCACATCGAACCGCGCGGCGGAAATCTACGTCAACGAAATCCACATCACCGGCA 1069

QY 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
|||||
Db 1070 ACAACAAACCCGCGAGGAGTGTGCGCGCGCAATTCGCCAAATGGAATCCGCGCCTT 1129

QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
|||||
Db 1130 ACGACACCTCAAGTCAACAGCTCCAAAGAGCGCGTCGAGCTTTTGGGCTACTTCGAC- 1188

QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
|||||
Db 1189 -----AACGTACAGATTGATGCGCTCCCGTCCCGGACACCGCACAAAGTCGATTGA 1243

QY 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
|||||
Db 1244 ACATGAGCCTGACCGAGCTTCCACCGCTCGCTCGAGCTTGAGCGCGGCTGGGTACAGG 1303

QY 310 luTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
|||||
Db 1304 ATACCGCCTGGTCATCTCCGCGAGCGCTTTCCCAAGACAACCTGTTCGGTACGGGC--- 1359

QY 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
|||||
Db 1360 --AAGTCGGCGCCCTGCGCGCCTCAGGAAGCAAAACCGCTCAACGCGCTCGCTGTCGT 1417

QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
|||||
Db 1418 TTACCGACCGTACTTCCAGCGCAGCGGGTCAAGCTGGGCTACGATGTTTACGGAAG 1477

QY 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
|||||
Db 1478 CCTTCGACCGCGCAAGCATCGACAGCATCAACAAATATATAAACCCACCGCAGCGC 1537

QY 373 euGluLysArgAlaPheSerGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
|||||
Db 1538 CAGGATCCCGATGAGCGTGCCTGTTACCGAATACGACCGCGTAATTCGGTTGGTGG 1597

QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
|||||
Db 1598 CAGAACCTTGACCGTCAACACCTACAAACAGCGCCCAACACTATGCGGACTTTATCA 1657

QY 401 laGluIArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
|||||
Db 1658 AGAAATACCGCAAAACCGACGGCACAGAC-----GCAGCTTCAAAAGGCTGGCTGT 1708

QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
|||||
Db 1709 ACAAGGTACCGTGGCGTGGGGCGCAACAAACCGACAGCGCGTTATGGCCGCGCGG 1768
```

```
QY 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
|||||
Db 1769 GC---TACCTGACGGGC-----GTGACGCCGCAAA 1795

QY 459 eulleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
|||||
Db 1796 TCGCCCTGCCCGCAGCAAACTGCAATACTACTCCGCCACCCCAACAAACCTGGTTCT 1855

QY 476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
|||||
Db 1856 TCCCTTAACCAAAACCTTCACGCTGATCTCGCGCGGAGTCCGCGGCGGCT 1915

QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSery 511
|||||
Db 1916 ACGCAGAACCAAAAGTATCCCTTCTTTGAAAACTTCTACGGCGGGCGCTGGTTCCG 1975

QY 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
|||||
Db 1976 TGGCGGATACGAAAGCGGACGCTCGGTCCGAAAGTGTATGACGAATACGGGCAAAAA 2035

QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
|||||
Db 2036 TCAGCTACGGCGGCAAC-----AAAAAGCCAAACGCTCCGCCGAGCTGC 2080

QY 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
|||||
Db 2081 TCTTCCCGATGCCCGCGCGCAAGACCGCGCAGCTCCGCTCAGGCTGTTTCCCGCAGC 2140

QY 557 etGlyAsp----- 559
|||||
Db 2141 CAGGCGCGTGTGGCGGCAAAACCTTACGACGACAAACAGAGTTCCGCGCAGCGCGCA 2200

QY 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
|||||
Db 2201 GGTTCAAAACATTACGCGCGCGCAATATCCCAATAATCCACCTTTACCAACGAATGC 2260

QY 570 yHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
|||||
Db 2261 GCTATTCCCGCGCGCGGCTTACCTGGCTCTCGCTTTAGGCCCGCATGAATTCAGCT 2320

QY 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
|||||
Db 2321 ACGCTACCGCTCGAAGAAAAACCGGAAGACGAAATCCCAACGCTTCCCAATTCCAACTCG 2380

QY 606 lyThrArgPhe 609
|||||
Db 2381 GCACGAGCTTC 2391

RESULT 39
AX202497
LOCUS AX202497 2394 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 10 from Patent WO0152885.
ACCESSION AX202497
VERSION AX202497.1 GI:15392217
KEYWORDS
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Plizza,M.C., Rappuoli,R.C. and Giuliani,M.C.
TITLE Outer membrane vesicle (omv) vaccine comprising N. meningitidis
JOURNAL serogroup b outer membrane proteins
Patent: WO 0152885-A 10 26-JUL-2001;
Chiron Spa (IT)
FEATURES
source 1..2394
Location/Qualifiers
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
Alignment Scores:
```

Pred. No.: 4.61e-08 Length: 2394
Score: 253.50 Matches: 151
Percent Similarity: 36.50% Conservative: 99
Best Local Similarity: 22.04% Mismatches: 276
Query Match: 7.99% Indels: 159
DB: 6 Gaps: 27

US-09-857-669-2 (1-609) x AX202497 (1-2394)

Qy 41 SerProaspThrGluSerValLysLeuLys-----ProLysPhe 53
Db 470 TCGCCCGCAACCGCGCTCCGACATCGACATGACGATGAGAGGGCAAAATCCCGCAAAATCA 529
Qy 54 ProVal-ArgIleAspThrGluIleAspSerGluIleLysAspMetValGluGluHisLeuPr 73
Db 530 CCGACATCGAATTCAGAGCAACCAAGCTATTCGACCGCAAACTGATCGCGGAGATGT 589
Qy 73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluInThrGlyPheLeuAlaGl 93
Db 590 CGCTGACCGAAGCGGCATTTGGACATGGCTCAGACGAAGCAACCAATTCACAGCAGA 649
Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe----- 109
Db 650 AATTTGCCCAAGATGGAAGAAAGTAACCGACTTCTACGACCAACACGCGTACTTCGATT 709
Qy 110 -----SerSerLysValSerLeuThrGluLysAspClyAlaTyrThrValHisI 126
Db 710 TCCGCATCCTCGATACCGACATCCCAACCAACGAAGCAACCAACCAACGACCATCAAA 769
Qy 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
Db 770 TCACCGTCCAGCAAGCGGAGCTTTCGCTGGGGCAAGTCTCCATC----- 816
Qy 144 spIleLeuSerAspClyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
Db 817 -----GAGGGCAGCACACCAACGAAGTCCCAAGCGCAACTGGAAAACTGTGTA 865
Qy 161 -----AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpG 176
Db 866 CCATGAAGCCCGCAATGGTACGAA-----C 892
Qy 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
Db 893 GCCAGCAGATCAGCGCGCTTTGGGTGAGATTCAGACCGCATGGCTCGCGCAGCGTACG 952
Qy 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
Db 953 CATACAGCGAAATC---AGCGTACAGCGCGCTGCCAAGCCGCAACCAACCAACCGCTCGATT 1009
Qy 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
Db 1010 TCGTCTCCATCGAAGCGCGCGGCAAAATCTAGTCAACGCAAAATCCACATCACCAGCA 1069
Qy 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
Db 1070 ACAACAAAACCGGACACCAAGCTTTCGCGCGCGAATTCGCGCAAAATGGAATCCGCGCGCTT 1129
Qy 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
Db 1130 ACGACACCTCAAGTGAACGCTCCCAAGAGCGCGTTCGAGCTTTTGGGCTACTTTCGAC- 1188
Qy 272 lyAlaSerValGlnAlaAspPheArgLeuGlnGly-----AspArgValProValL 290
Db 1189 -----AAGGTACAGTTTCATCGCTCGCGCTTGGCGGCACACCGACAAAGTCGATTGGA 1243
Qy 290 ysvAlSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
Db 1244 ACATGAGCGCTACCGCAAGCTTTCACCGCGCTCGCTCGACTTTCAGCGCGCGGCTGAGTACAGG 1303
Qy 310 luTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
Db 1304 ATACCGGGCTGGTATGTCCCGCAGCGGTTTCCCAAGACACACTGTTCGGTACGGGC----- 1359

Qy 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
Db 1360 --AAGTCGGCGCGCTCGCGCGCTCACGAAGCAAAACACCACTCAACGGCTCGTGTGCGT 1417
Qy 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgLysAspT 358
Db 1418 TTACCGACCGCTACTTTCACGCGACAGCGGCTCAGCGTGGCTGAGTGTTCACGGAAGA 1477
Qy 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
Db 1478 CTTTCGACCGCGCAAGACATCGACGACCAACCAATATAAACCAACCAACCGCAGCGG 1537
Qy 373 euGluLysArgAlaPheSerGlyGlyIleTyrValArgAspArgAlaGlyIleAspA 393
Db 1538 CAGGCATCCCGATCAGCGTGGCTTACCGAATACGACCGCGTCAATTCGGTTTGGTGG 1597
Qy 393 la-----ArgLeuGlyAlaGluPheLeuA 401
Db 1598 CAGAACACCTGACCGTCAACACCTTACAACAAAGCGCCCAACACACTATGCCGACTTTATCA 1657
Qy 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
Db 1658 AGAAATACGGCAACACCGACGCGACAGC-----GGCAGCTTCAAGGCTGGGTGT 1708
Qy 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnG 439
Db 1709 ACAAGGTACCGTGGCTGGGGCGCAACAAACGACAGCGCGTATTCGCGCAGCGCG 1768
Qy 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
Db 1769 GC---TACCTGACCGGC-----GTGAACCGCGAAA 1795
Qy 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
Db 1796 TCGCGCTCGCGCGCAGCAAACTGCAATACTACTCCGCCACCCACCAACCAACCTGGTTCT 1855
Qy 476 -----LysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaA 492
Db 1856 TCCCTTTAAGCAAAACCTTCACGCTGATGCTCGCGCGCAAGTCGCGATTTCGGCGGCT 1915
Qy 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerV 511
Db 1916 ACGCAGACCAACCAAGAAATCCCTTCTTGAATACTTCTACGGCGCGCGCTGGGTTCGG 1975
Qy 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Db 1976 TCGCGGATACGAAGCGCGCAGCTCGGTCCGAAAGTGTATGACGAATACGCGCAAAAAA 2035
Qy 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
Db 2036 TCAGCTACGGCGGCAAC-----AAAAAGCCACGCTCTCCCGCGAGGTGC 2080
Qy 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Db 2081 TCTTCCCGATCCCGCGCGCAAGACGCGCGCTCGCGCTCGCGCTGTTCGCCAGC 2140
Qy 557 etGlyAsp----- 559
Db 2141 CAGCAGCGGTGGCAGCGCAAAACCTACGACGACCAACAGCAGTTCCGCGCAGCGCGCA 2200
Qy 560 -----AlaAlaAlaAsnPhelLysArgMet-----LysLeuL 570
Db 2201 GGGTTCAAAACATTTACGGCGCGCGCAATACCCATTAATCCACCTTTACCAACCAATTCG 2260
Qy 570 ysvHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 2261 GCTATTCCGCGCGCGCGGCTTACCTGGCTCTCGCGCTTAGCGCGCGATGAATTCAGCT 2320
Qy 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpIleHisSerLeuG 606
Db 2321 ACGCTACCGCTCAGANAAAAACCGGNAGACGAATAACACGCTTCCCAATTCCAACTCG 2380
Qy 606 lyThrArgPhe 609

Db	2381	GCACGACGTT	2391	
RESULT 40				
LOCUS	NMA122491	340806 bp	DNA linear	BCT 04-DEC-2000
DEFINITION	Neisseria meningitidis serogroup A strain 22491 complete genome;			
ACCESSION	AL162752	AL157959		
VERSION	AL162752.2	GI:7378778		
KEYWORDS				
SOURCE	Neisseria meningitidis 22491.			
ORGANISM	Neisseria meningitidis 22491			
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;			
AUTHORS	Neisseria.			
	1 (bases 1 to 340806)			
	Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,			
	Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,			
	Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,			
	Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,			
	Quail,M.A., Rajadream,M.A., Rutherford,K.M., Simmonds,M.,			
	Skelton,J., Whitehead,S., Spratt,B.G. and Barrrell,B.G.			
TITLE	Complete DNA sequence of a serogroup A strain of Neisseria			
JOURNAL	Meningitidis 22491			
MEDLINE	Nature 404 (6777), 502-506 (2000)			
PUBMED	20222536			
REFERENCE	10761919			
AUTHORS	2 (bases 1 to 340806)			
	Parkhill,J.			
TITLE	Direct Submission			
JOURNAL	Submitted (30-MAR-2000) Submitted on behalf of the Neisseria			
	sequencing team, Sanger Centre, Wellcome Trust Genome Campus,			
	Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk			
COMMENT	Notes:			
	Details of N. meningitidis sequencing at the Sanger Centre are			
	available on the World Wide Web.			
	(URL, http://www.sanger.ac.uk/projects/N_meningitidis/).			
FEATURES	Location/Qualifiers			
source	1..340806			
	/organism="Neisseria meningitidis 22491"			
	/strain="22491"			
	/db_xref="taxon:122587"			
gene	/note="serogroup: A"			
	complement(38..358)			
	/gene="NMA0001"			
	complement(38..358)			
CDS	/gene="NMA0001"			
	/note="NMA0001, unknown, len: 106 aa"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical protein NMA0001"			
	/protein_id="CAB83321.1"			
	/db_xref="GI:7378779"			
	/db_xref="SPTREMBL:Q9JX93"			
	/translation="MSNQTIRSDFTTSRAWGALDIANNMGTTVRLHWTDPQYKWHV			
	NDGEVFVAMDGEVDMHYRENGEHHIVRLKSGDIFYAGIGTGHVHPRGETRILVIEK			
	EGSV"			
	complement(414..2438)			
gene	/gene="nuoL"			
	complement(414..2438)			
CDS	/gene="nuoL"			
	/EC_number="1.6.5.3"			
	/note="NMA0002, nuoL, NADH dehydrogenase I chain L, len:			
	674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase			
	I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to			
	entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone			
	(complex I), various chains, and to entry PF00662			
	oxidored_q1_N, NADH-Ubiquinone oxidoreductase (complex I),			
	chain 5 N-terminus"			
	/codon_start			

```

/EC_number="1.6.5.3"
/Note="NMA0005, nuok, NADH dehydrogenase I chain K, len:
101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase
I chain K (EC 1.6.5.3) (102 aa), fasta scores: E():
1.le-24: 65.0% identity in 100 aa overlap. Contains Pfam
match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain K"
/protein_id="CAB83325.1"
/db_xref="GI:7378783"
/db_xref="SPTREMBL:Q9JQ09"
/transl_table="MITLHYLVIGALLFGISAMGIFPMNRKNVLVLLMSIELMLLAYN
FNFTAFSQHLDGTAGQIFVFVLTVAASAIGLAIMLVYRNQTINVADLDELKG"
/complement(3564..3863)
/gene="nuok"
/Note="Pfam match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
score 141.50, E-value 1.5e-38"
/complement(3863..4534)
/gene="nuoJ"
/complement(3863..4534)
/gene="nuoJ"
/EC_number="1.6.5.3"
/Note="NMA0006, nuoj, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fasta scores: E():
9.7e-21: 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/transl_table="MTFOLILFYFAVILLYGAIKTVTKPNVHAALHLVITFCVSAM
LMMLQAEFLGTVLVVYGVAVMVLFLVVMVLMNIDIEEMRAGFWRHAPVAVVGTLL
AVALLILVNPKTDLAAFLGMLKIDPADYNINRDLGSRIYTDYLLPFLAALVLLLGMY
AATLVHRKTVNPKRMDPADQVKVADQGRMLVKMEAVRPQTESAESEVSDDLKPK
EEGKA"
/complement(3874..3879)
/gene="nuoJ"
/complement(4034..4525)
/gene="nuoJ"
/Note="Pfam match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"
/gene="NMA0007"
/complement(4570..5277)
/gene="NMA0007"
/Note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"

```

Alignment Scores:

```

Pred. No.:      3.09e-05      Length:      340806
Score:          253.50      Matches:      151
Percent Similarity: 36.50%      Conservative: 99
Best Local Similarity: 22.04%      Mismatches: 276
Query Match:      7.99%      Indels:      159
DB:              1          Gaps:      27

```

US-09-857-669-2 (1-609) x NMA12491 (1-340806)

```

Oy  41 SerProAspThrGluserVallylsLeuLys-----ProLysPhe 53
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84451 TCGCCCGGACCGGCTCGACATCGACATGACGATGACGGGCAATCCGCCAATCA 84510

```

```

Oy  54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHIsLeuPr 73
|||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84511 CCGACATCGAATTTGAAGGCAACCAAGTCTATTCGACCGCAAACTGATCGGCGAGATGT 84570

Oy  73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAgl 93
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  84571 CGCTGACCGAAGCGGCATTTGGACATGGCTCAGACGAAACCAATTCACAGCAGAGA 84630

Oy  93 uLluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe---- 109
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84631 AATTTGCCACAGACATGGAATAAGTAACCGACTTCTACGACGAACACCGCTACTTCGATT 84690

Oy  110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHIsI 126
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84691 TCCGCATCTCTGCATACCATCCAAACCAACGAAGACAAACCAACGACGACCATCAAAA 84750

Oy  126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  84751 TCACCGTCCAGGAGCGGAGCGTTTCCGTTGGGGCAAAAGTCTCCATC----- 84797

Oy  144 spIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  84798 -----GAAGGCGACACCAACGAAGTCCCAAGCGCAACTGGAAAAAATCGCTGA 84846

Oy  161 -----AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpG 176
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84847 CCATGAACCGCGGCAATCGTACGAA-----C 84873

Oy  176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  84874 GCCAGCAGATGACCGCGCTTTTGGTGAGATTTCAGAACGCGATGGCGCTCGGCGAGCTACG 84933

Oy  192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  84934 CATACAGCGAAATC---AGCGTACAGCGCGTCCCAACCGCGCAACCAACCGCTCGATT 84990

Oy  212 euAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  84991 TCGTCTCGACATCGAACCGCGCGGCAAAATCTAGCTCAACGAATCCACATCACCGCA 85050

Oy  232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85051 ACAACAAACCGCGCGAGAGTCTGTCGCGCGCAATTCGCGCAATGGAATCGCGCGCTT 85110

Oy  252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHIsTyrSerG 272
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  85111 ACGACACCTCCAGTGCACGCTCCAAAGAGCGCGTCTGGCTTTGGGCTACTTCGAC- 85169

Oy  272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85170 -----AACGTACAGTTTCATGCGGTCGCGCTTGGCGGCACACCGCACAAAGTCGATTGA 85224

Oy  290 ysValSerValThrGluValLysArgHIsLysLeuGluThrGlyIleArgLeuAspSerG 310
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  85225 ACATGAGCTGACCGAGACGCTCCACCGCGCTCGCTAGCTTCAGCGCGGCGTGGGTACAGG 85284

Oy  310 luTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85285 ATACCGGCTGTGTCATGTCGCGAGCGGCTTCCCAAGACACACTGTTCGGTACGGGC---- 85340

Oy  330 leGlySerValValTrpAspMetAspLysTyrGluThrLeu----- 344
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85341 --AAGTCGCGCGCGCTCGCGCTCCACGAAGCAAAACACGCTCAACGCGCTCGGTGCGT 85398

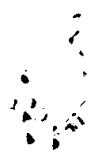
Oy  345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85399 TTACCGACCGCTACTTTCACGCGACAGCGGCTCGGCTGACGATTTTACGGAAGAAG 85458

Oy  358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  85459 CTTTCGACCGCGGCAACGATCGACGACATCAACATATNTAATACCAACACCGCGGCGG 85518

```

```
QY 373 euGlulysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
Db 85519 CAGGCATCGCATGAGCGTGCCTGTACCGAATACGACGCGTGAATTCCTGGTTTGGTGG 85578
QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
Db 85579 CAGAACACCTGACCGTCAACACCTACAAAGCGGCCCAACACACTATGCCGACTTTATCA 85638
QY 401 ladlulGlyArgGlyIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
Db 85639 AGAATACCGCAAAACCGACGCGCACAGAC-----GGCAGCTTCAAGGCTGCCTGT 85689
QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
Db 85690 ACAAGGTACCGTGGCTGGGGCGCAACAAACACGACAGCGGTTATGGCCGACGCGCG 85749
QY 439 lyHisTrpLeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaL 459
Db 85750 GC---TACCTGAGGGC-----GTCAACGCGCGAAA 85776
QY 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
Db 85777 TCGCCCTGCCGCGCAGCAAACTGCAATACTACTCCGCCACCCACAAACCACTGGTTCT 85836
QY 476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
Db 85837 TCCCTTTAAGCAAAACCTTCACGCTGATCTCGCGCGGAAGTCGSCATTGGCGGCGGT 85896
QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerServ 511
Db 85897 ACGGCAGAACCAAGAAATCCCTCTTTGAAAACTTCTACGGCGCGCGCTGGGTTCCG 85956
QY 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Db 85957 TCGCGCGATACGAAACGCGCAGCTCGTCCGNAAGTGTATACGCAATACGGCGGAAAAA 86016
QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuG 540
Db 86017 TCAGCTACGGCGGCAAC-----AAAAAGCCAAACGTCTCCGCGAGCTGC 86061
QY 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Db 86062 TCTTCCCGATGCCGCGCGGCAAGACGCGCACCTCGCTGAGCCTGTTTGGCGAGG 86121
QY 557 etGlyAsp----- 559
Db 86122 CAGGCAGCGTGTGGGAGCGCAAAACCTACGACGACAAACAGCAGTTCCTCCGCGCGCGGCA 86181
QY 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
Db 86182 GGGTTCAAAACATTTACGGCGCGGCAATACCCATAATCCACCTTTTACCACGAATTGC 86241
QY 570 yshiclySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 86242 GCTATTCCCGCGCGCGGCGGTACCTGCTCTCGCTTTAGGCGCGATGAAATTCAGCT 86301
QY 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
Db 86302 ACGCTTACCGGTGAAGAAAAACCGGAAGACGAATAATCCAAACGCTTCCAATTCCAACTCG 86361
QY 606 lyThrArgPhe 609
Db 86362 GCACGAGGTTTC 86372
```

Search completed: November 9, 2002, 02:46:21
Job time : 4583 secs



GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: November 9, 2002, 01:03:13 ; Search time 36 Seconds
(without alignments)
2254.156 Million cell updates/sec
Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLGTRF 609
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 908470 seqs, 133250620 residues
Word size : 15
Total number of hits satisfying chosen parameters: 6
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 1000 summaries

Database : A_Geneseq_l01002:*				
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*			
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*			
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*			
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*			
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*			
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*			
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*			
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*			
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*			
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*			
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*			
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*			
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*			
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*			
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*			
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*			
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*			
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*			
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*			
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*			
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*			
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*			
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	609	100.0	609 21	AA199623
2	609	100.0	609 21	AA199624
3	249	40.9	615 21	AA194862
4	198	32.5	587 21	AA199625
5	198	32.5	615 21	AA194861
6	170	27.9	522 21	AA194860

ALIGNMENTS

RESULT 1	
AA199623	
ID	AA199623 standard; Protein; 609 AA.
XX	
AC	AA199623;
XX	
DT	03-JAN-2001 (first entry)
XX	
DE	Neisseria meningitidis BASB040 putative protein sequence #1.
XX	
BASB040;	bacterial disease; respiratory tract infection; bacteraemia; meningitis; cancer; autoimmune disease.
KW	
XX	Neisseria meningitidis.
OS	
XX	
PN	WO200034480-A1.
XX	
PD	15-JUN-2000.
XX	
PF	02-DEC-1999; 99WO-EP09560.
XX	
PR	07-DEC-1998; 98GB-0026886.
XX	
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX	
PI	Ruelle J;
XX	
DR	WPI; 2000-423436/36.
DR	N-PSDB; AAA48507.
XX	
PT	Novel BASB040 polypeptides of Neisseria meningitidis useful for diagnostic, prophylactic and therapeutic purposes against microbial diseases comprise a specific amino acid sequence -
PT	
XX	
PS	Claim 4; Page 59-60; 98pp: English.
XX	
CC	The present sequence is a putative version of the Neisseria meningitidis strain ATCC13090 BASB040 protein sequence. This protein is similar to the D15 outer membrane protein of the bacterium. The protein, its gene, antibodies, antagonists and agonists can be used to diagnose and treat bacterial diseases such as those leading to upper respiratory tract infections, bacteraemia and meningitis. In addition, they can be used in vaccines for use against cancer and autoimmune diseases.
XX	
SQ	Sequence 609 AA;
Query Match 100.0%; Score 609; DB 21; Length 609;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MMKPTALLPALFFPHAYAPADLSNKAAGFALFNKSPDTSVKLKPFPVRIDTQ 60
Db	1 MMKPTALLPALFFPHAYAPADLSNKAAGFALFNKSPDTSVKLKPFPVRIDTQ 60
QY	61 DSEIKDVEEHLPLITQOQEEVLDEKQETGFLAEAPDNVKTMLRSKGVFSSKSLTEKDG 120
Db	61 DSEIKDVEEHLPLITQOQEEVLDEKQETGFLAEAPDNVKTMLRSKGVFSSKSLTEKDG 120
QY	121 AYTTHITPGPRTKTIANVGVAITGLDSDGNLAEYRNALENWQOPVGSDFDQDSWENSKT 180
Db	121 AYTTHITPGPRTKTIANVGVAITGLDSDGNLAEYRNALENWQOPVGSDFDQDSWENSKT 180
QY	181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db	181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVVDSGRPIAFGDFEITGTQRYPEQIV 240
QY	241 SGLARFOPGTPYDLDLLDFOALEQNGHYSGASVQADFDRLOGDRVPKVSVEVKRHK 300
Db	241 SGLARFOPGTPYDLDLLDFOALEQNGHYSGASVQADFDRLOGDRVPKVSVEVKRHK 300
QY	301 LETGIRLDSYGLGCKIAYDYNNLFNKGYIGSVVYMDMDKYETTLAAGISOPRNYRGNYWT 360

Db 301 LETGIRLDSEYGLGKIAIDYNNLFNKGYIGSVVWMDKYEITTLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTTQNLKRAFSGGIWYVRDAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTTQNLKRAFSGGIWYVRDAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Qy 481 IIRQAGYTVARDNADVPGLMFRSGGASSVRYGVELDSIGLAGPNSVLPERALLVGSLE 540
Db 481 IIRQAGYTVARDNADVPGLMFRSGGASSVRYGVELDSIGLAGPNSVLPERALLVGSLE 540
Qy 541 YQLPFTRTLSCAVFDMGDAANFKRMKLKHGSGLGVRWFSPAPFSFDIAYGHSKDKIR 600
Db 541 YQLPFTRTLSCAVFDMGDAANFKRMKLKHGSGLGVRWFSPAPFSFDIAYGHSKDKIR 600
Qy 601 WHISLGTRF 609
Db 601 WHISLGTRF 609

RESULT 2

AAY99624
ID AAY99624 standard; Protein: 609 AA.

XX AC AAY99624;

XX DT 03-JAN-2001 (first entry)

XX DE Neisseria meningitidis BASB040 putative protein sequence #2.

XX KW BASB040; bacterial disease: respiratory tract infection; bacteraemia;

XX KW meningitis; cancer; autoimmune disease.

XX OS Neisseria meningitidis.

XX PN WO200034480-A1.

XX PD 15-JUN-2000.

XX PF 02-DEC-1999; 99WO-EP09560.

XX PR 07-DEC-1998; 98GB-0026886.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX XX WPI; 2000-423426/36.

DR N-PSDB; AAA48508.

XX PT Novel BASB040 polypeptides of Neisseria meningitidis useful for

PT diagnostic, prophylactic and therapeutic purposes against microbial

XX diseases comprise a specific amino acid sequence

XX PS Claim 4; Page 60-61; 98pp; English.

XX CC The present sequence is a putative version of the Neisseria

CC meningitidis strain ATCC13090 BASB040 protein sequence. This protein is

CC similar to the D15 outer membrane protein of the bacterium. The protein,

CC its gene, antibodies, antagonists and agonists can be used to diagnose

CC and treat bacterial diseases such as those leading to upper respiratory

CC tract infections, bacteraemia and meningitis. In addition, they can be

XX used in vaccines for use against cancer and autoimmune diseases.

XX Sequence 609 AA;

Query Match 100.0%; Score 609; DB 21; Length 609;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMIKPTALLLPALPFFFFPHAYAPAAADLSENKAAGFALPKNKSPTDESVKLKPKFPVRIIDTQ 60
Db 1 MMIKPTALLLPALPFFFFPHAYAPAAADLSENKAAGFALPKNKSPTDESVKLKPKFPVRIIDTQ 60
Qy 61 DSEIKDWVEEHLPLITOOQEEVLQKEQTGFLEAEAPDNVKTMLRSKGYFSSKVSILTEKDG 120
Db 61 DSEIKDWVEEHLPLITOOQEEVLQKEQTGFLEAEAPDNVKTMLRSKGYFSSKVSILTEKDG 120
Qy 121 AYTWHITPGPRTKIANVGVAILGDI LSDGNLAEEYRYNALENWQOPVGSDFDQDQSWENSKT 180
Db 121 AYTWHITPGPRTKIANVGVAILGDI LSDGNLAEEYRYNALENWQOPVGSDFDQDQSWENSKT 180
Qy 181 SVLGAVTRKGYPLAKLGNTRAANVPDTATVDLNVVDSGRP IAFGDPEITCTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAANVPDTATVDLNVVDSGRP IAFGDPEITCTQRYPEQIV 240
Qy 241 SGLARFQPGTPTDLDLLDFQOALEQNGHYSYGASVQADFRLQGDVVPVKYSVTEVKRHK 300
Db 241 SGLARFQPGTPTDLDLLDFQOALEQNGHYSYGASVQADFRLQGDVVPVKYSVTEVKRHK 300
Qy 301 LETGIRLDSEYGLGKIAIDYNNLFNKGYIGSVVWMDKYEITTLAAGISQPRNRYGNWT 360
Db 301 LETGIRLDSEYGLGKIAIDYNNLFNKGYIGSVVWMDKYEITTLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTTQNLKRAFSGGIWYVRDAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTTQNLKRAFSGGIWYVRDAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Qy 481 IIRQAGYTVARDNADVPGLMFRSGGASSVRYGVELDSIGLAGPNSVLPERALLVGSLE 540
Db 481 IIRQAGYTVARDNADVPGLMFRSGGASSVRYGVELDSIGLAGPNSVLPERALLVGSLE 540
Qy 541 YQLPFTRTLSCAVFDMGDAANFKRMKLKHGSGLGVRWFSPAPFSFDIAYGHSKDKIR 600
Db 541 YQLPFTRTLSCAVFDMGDAANFKRMKLKHGSGLGVRWFSPAPFSFDIAYGHSKDKIR 600
Qy 601 WHISLGTRF 609
Db 601 WHISLGTRF 609
RESULT 3
AAY74862
ID AAY74862 standard; Protein: 615 AA.
XX AC AAY74862;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1198.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 03-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.

DE	Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1196.
XX	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW	antibacterial; gene therapy.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO9957280-A2.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99WO-US09346.
XX	
PR	01-MAY-1998; 98US-0083758.
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0098994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	
PA	{CHIR } CHIRON CORP.
PA	{GENO-} INST GENOMIC RES.
XX	
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI	Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI	Tettelin H, Venter JC;
DR	WPI: 2000-062150/05.
DR	N-PSDB; AAZ53623.
XX	
PT	Novel Neisserial polypeptides predicted to be useful antigens for
PT	vaccines and diagnostics -
XX	
PS	Claim 2; Page 667; 1453pp; English.
XX	
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC	represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polynucleotides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunogenic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC	presence of Neisseria bacteria, or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polynucleotides of the invention
CC	may also be used in gene therapy protocols.
XX	
SQ	Sequence 615 AA;
	Query Match 32.5%; Score 198; DB 21; Length 615;
	Best Local Similarity 100.0%; Pred. No. 3.4e-192;
	Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	412 DLGNSHATMLTASWKROLLNNVLHPENGHYLDGIGTTGLTFSLSTALIRTSARAGYEFT 471
Db	418 DLGNSHATMLTASWKROLLNNVLHPENGHYLDGIGTTGLTFSLSTALIRTSARAGYEFT 477
OY	472 PENKKLGTFFIIRGOAGVTYARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNGSVLP 531
Db	478 PENKKLGTFFIIRGOAGVTYARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNGSVLP 537
OY	532 RALLVGSLEYQLPTRLTSLCAVFHDMGDAAANFRMKLKHGSGLCGVRFWSPLAFPSFDIA 591
Db	538 RALLVGSLEYQLPTRLTSLCAVFHDMGDAAANFRMKLKHGSGLCGVRFWSPLAFPSFDIA 597
OY	592 YGHSDDKIRWHISLGTRF 609
Db	598 YGHSDDKIRWHISLGTRF 615

|||||
Db 106 KTLRSKGYESSKVSLETKDGYTVHITPGPRTKIANVGVAIILGDILSDGNLAETYNAL 165
QY 160 ENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKIGNTRAANPDAT 209
Db 166 ENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKIGNTRAANPDAT 215

Search completed: November 9, 2002, 01:11:27
Job time : 36 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:10:43 ; Search time 20 Seconds
(without alignments)
895.928 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMKPTALLLPALFFPPHAY.....IAYGSDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID

Description					

No matches found

Search completed: November 9, 2002, 01:14:22
Job time : 20 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:13:03 ; Search time 14 Seconds
(without alignments)
627.205 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMKPTALLLPALFFPPHAY.....IAYGHSDDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 92612 seqs, 14418503 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
---------------	----------------	--------	----	-------------

No matches found

Search completed: November 9, 2002, 01:17:39
Job time : 14 secs

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:10:03 ; Search time 21 seconds
(without alignments)
2787.898 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 609

Sequence: 1 MMIKPTALLLPALFFPPHAY.....IAYGHSDKKIRNHISLGTFR 609

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	249	40.9	615 2 A82025	probable outer mem
2	198	32.5	635 2 G81003	conserved hypothet

ALIGNMENTS

RESULT 1

A82025

probable outer membrane protein NMA0296 [imported] - Neisseria meningitidis (strain 2249

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-feb-2001

C:Accession: A82025

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel ; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: A82025

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-615 <PAR>

A:Cross-references: GB:AL157959; NID:g7378778; PIDN:CAB83602.1; PID:g737905

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: NMA0296

Query Match 40.9%; Score 249; DB 2; Length 615;

Best Local Similarity 99.6%; Pred. No. 8.3e-249;

Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1	MMIKPTALLLPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ	60	
Db	7	MMIKPTALLLPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ	66	
Qy	61	DSEIKDMVEEHLPLITQQQEEVLDEKQGTGFLAEEAPDNVKTMLRSKGYFSSKVSUTEKDG	120	
Db	67	DSEIKDMVEEHLPLITQQQEEVLDEKQGTGFLAEEAPDNVKTMLRSKGYFSSKVSUTEKDG	126	
Qy	121	AYTVHIITPGPRTKIANVGVAIIGDILSDGNLAETRYNRALENWQOPVGSDFDQDQSWENSKT	180	
Db	127	AYTVHIITPGPRTKIANVGVAIIGDILSDGNLAETRYNRALENWQOPVGSDFDQDQSWENSKT	186	
Qy	181	SVLGAVTRKGYPLAKLGNTRAAVNPDATVDLNVVDSGRPIAFGDFRITCTQRYPEQIV	240	
Db	187	SVLGAVTRKGYPLAKLGNTRAAVNPDATVDLNVVDSGRPIAFGDFRITCTQRYPEQIV	246	
Qy	241	SGLARFQPGTGYDLDDLLDFQALQNGHYSGASVQADFDRLQGDVRVPKYSVTEVKRHK	300	
Db	247	SGLARFQPGTGYDLDDLLDFQALQNGHYSGASVQADFDRLQGDVRVPKYSVTEVKRHK	306	
Qy	301	LETGIRLDSEYGLGKTIAYDYNNFNKGYIGSVVMDMKYETTLAAGISQPRNRYGNWT	360	
Db	307	LETGIRLDSEYGLGKTIAYDYNNFNKGYIGSVVMDMKYETTLAAGISQPRNRYGNWT	366	
Qy	361	SNVSYNRSTTONLEKRAFSGGIWYVRDAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM	420	
Db	367	SNVSYNRSTTONLEKRAFSGGIWYVRDAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM	426	
Qy	421	LTASWKROLNNVLHPENGHYLDGKIGTTLG	451	
Db	427	LTASWKROLNNVLHPENGHYLDGKIGTTLG	457	

RESULT 2

G81003

conserved hypothetical protein NMB2134 [imported] - Neisseria meningitidis (strain

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: G81003

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: G81003

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-635 <TET>

A:Cross-references: GB:AF002561; GB:AB002098; NID:g7227384; PIDN:AAP42442.1; PID:g72

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB2134

Query Match 32.5%; Score 198; DB 2; Length 635;

Best Local Similarity 100.0%; Pred. No. 5.1e-196;

Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 DLGNSHATMLTASWKROLNNVLHPENGHYLDGKIGTTLGTLSTALIRTSARAGYFFT 471

Db 438 DLGNSHATMLTASWKROLNNVLHPENGHYLDGKIGTTLGTLSTALIRTSARAGYFFT 497

Qy 472 PENKKLGTFIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 531

Db 498 PENKKLGTFIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 557

Qy 532 RALLVGSLEYOLPFTRTLTSAGVFDHMDGAAANFKRMKLKHGSGLVGRWFSPLAFSFDIA 591

Db 558 RALLVGSLEYOLPFTRTLTSAGVFDHMDGAAANFKRMKLKHGSGLVGRWFSPLAFSFDIA 617

Qy 592 YGHSDDKIRWHISLGRF 609
| | | | | | | | | | | | | | | |
Db 618 YGHSDDKIRWHISLGRF 635

Search completed: November 9, 2002, 01:13:41
Job time : 21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:03:58 ; Search time 17 Seconds
(without alignments)
1485.828 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMKPTALLLPALFFPHAY.....IAYGHSDDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: November 9, 2002, 01:12:05
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:09:24 ; Search time 33 seconds
(without alignments)
3802.506 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMIKPTALLPALFFPPHAY.....IAYGHSDKKIRWHISLGRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 15
Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_nhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	249	40.9	615	16 Q9JWM5	Q9jwm5 neisseria m
2	198	32.5	635	16 Q9JXB7	Q9jxb7 neisseria m

SUMMARIES

ALIGNMENTS

RESULT 1
Q9JWM5 ID Q9JWM5 PRELIMINARY; PRT; 615 AA.
AC Q9JWM5;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE Putative outer membrane protein.
GN NMA0296.

RESULT 2
Q9JXB7 ID Q9JXB7 PRELIMINARY; PRT; 635 AA.
AC Q9JXB7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.

OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mc58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;

OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leath S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
RT *Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.*
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83602.1; -.
KW Complete proteome.
SQ SEQUENCE 615 AA; 67775 MW; B9A64607638E1CCB CRC64;
Query Match 40.9%; Score 249; DB 16; Length 615;
Best Local Similarity 99.6%; Pred No. 2.6e-249;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MMIKPTALLPALFFPPHAYAAADLSNKAAGFALFKNKSPTDESVKLKPFPVRI DTQ 60
Db 7 MMIKPTALLPALFFPPHAYAAADLSNKAAGFALFKNKSPTDESVKLKPFPVRI DTQ 66
Qy 61 DSEIKDMVEEHLPLITQOOEEVLKKEQTGFLAEAPDNVKTMLRSKGYFSSKVS LTEKDG 120
Db 67 DSEIKDMVEEHLPLITQOOEEVLKKEQTGFLAEAPDNVKTMLRSKGYFSSKVS LTEKDG 126
Qy 121 AYTWHITPGPRTKIANVGVAITGLDILSDGNLAEYRNALNNQOPVGSDFDQDSWENSKT 180
Db 127 AYTWHITPGPRTKIANVGVAITGLDILSDGNLAEYRNALNNQOPVGSDFDQDSWENSKT 186
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITCTQRYPEQIV 240
Db 187 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITCTQRYPEQIV 246
Qy 241 SGLARFQPGTPTDLDLDDFOQALEQNGHYSGASVQADFRLQGRVPVKVSVTEVKRHK 300
Db 247 SGLARFQPGTPTDLDLDDFOQALEQNGHYSGASVQADFRLQGRVPVKVSVTEVKRHK 306
Qy 301 LETGIRLDSEYGLGKTAIDYINFLNFKYIGSVVMDMDKYETTLAAGISQPRNYRGNYWT 360
Db 307 LETGIRLDSEYGLGKTAIDYINFLNFKYIGSVVMDMDKYETTLAAGISQPRNYRGNYWT 366
Qy 361 SNVSYNRSTTQNLKRAFSGGIWYVRDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSYNRSTTQNLKRAFSGGIWYVRDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 426
Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTTLG 451
Db 427 LTASWKROLLNNVLPENGHYLDGKIGTTLG 457
RESULT 2
Q9JXB7 ID Q9JXB7 PRELIMINARY; PRT; 635 AA.
AC Q9JXB7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.

```
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MCS8.*;
RL Science 287:1809-1815(2000).
RL EMBL: AE002561; AAF4242.1; -.
DR TIGR; NMB2134; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA; 69801 MW; 2B371F4A7A9DDA98 CRC64;

Query Match      32.58; Score 198; DB 16; Length 635;
Best Local Similarity 100.0%; Pred: No. 2.5e-196;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 DLGNSHATMLTASWKROLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFT 471
   |||||||
Db 438 DLGNSHATMLTASWKROLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFT 497
   |||||||

Qy 472 PENKKLGTFIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSTGLAGPNSVLP 531
   |||||||
Db 498 PENKKLGTFIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSTGLAGPNSVLP 557
   |||||||

Qy 532 RALLVGSLEYOLPPTRTLSGAVFHDMDGAAANFKRMKLGSGGLGVRFSPPLAPFSDIA 591
   |||||||
Db 558 RALLVGSLEYOLPPTRTLSGAVFHDMDGAAANFKRMKLGSGGLGVRFSPPLAPFSDIA 617
   |||||||

Qy 592 YGHSDDKIRWHISLGRF 609
   |||||||
Db 618 YGHSDDKIRWHISLGRF 635
```

Search completed: November 9, 2002, 01:12:59
Job time : 33 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 01:21:04 ; Search time 266 seconds
(without alignments)
5155.891 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLGRF 609

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgcn2_1/USPTO.spool/US09857669/runat_05112002_110007_7714/app_query.fasta_1.775
-DB=N-Geneseq.101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669 -CGCN_1_1_145 -runat_05112002_110007_7714 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N-Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3173	100.0	1830	21	AAA48507	Neisseria meningit
2	3173	100.0	1830	21	AAA48508	Neisseria meningit
3	3153	99.4	1848	21	AAZ53624	Neisseria meningit
4	3153	98.8	1848	21	AAZ53623	Neisseria meningit
5	3136	98.8	92934	21	AA841473	N. meningitidis pa
6	3136	98.8	172325	21	AAF21613	Neisseria meningit
7	3019	95.1	1764	21	AAA48509	Neisseria meningit
8	2653	83.6	1569	21	AAZ53622	Neisseria gonorrhoe
9	384.5	12.1	1374	22	AAF61044	P. putida KT2440-a
10	374	11.8	1737	21	AAA50269	Haemophilus influe
11	355.5	11.2	1731	21	AAA50270	Haemophilus influe
12	260.5	8.2	2505	21	AAA15155	DNA encoding outer
13	256.5	8.1	2379	22	AAH07278	Neisseria meningit
14	256.5	8.1	2379	22	AAH42129	Nucleotide sequenc
15	253.5	8.0	2394	22	AAH07279	Neisseria gonorrhoe
16	253.5	8.0	2394	22	AAH42130	Nucleotide sequenc
17	253	8.0	2394	22	AAH07277	Neisseria meningit
18	253	8.0	2394	22	AAH42128	Nucleotide sequenc
19	253	8.0	52253	21	AA81478	N. meningitidis pa
20	253	8.0	349980	21	AAF21544	Neisseria meningit
21	251.5	7.9	2394	21	AA15156	DNA encoding outer
22	247	7.8	6617	22	AA197964	Lawsonia intracell
23	241	7.6	2193	23	AA593089	DNA encoding novel
24	234.5	7.4	2388	24	ABQ72980	Methylococcus caps
25	234.5	7.4	2391	24	ABQ90024	M. capsulatus gene
26	231.5	7.3	1425	23	AA581897	DNA encoding novel
27	231.5	7.3	1425	23	AA593087	DNA encoding novel
28	219	6.9	2670	21	AAA50537	Moraxella catarrha
29	219	6.9	2760	21	AAA50536	Moraxella catarrha
30	219	6.9	89047	22	AAF28547	Genomic fragment #
31	213.5	6.7	2989	15	AAQ66202	H. influenzae PAK
32	212.5	6.7	944	24	ABK52155	Haemophilus influe
33	210.5	6.6	2974	15	AAQ66201	H. influenzae SB33
34	209.5	6.6	2984	15	AAQ66199	H. influenzae b Ea
35	208.5	6.6	2953	15	AAQ66200	H. influenzae b Mi
36	207.5	6.5	2949	15	AAQ66198	H. influenzae b Ca
37	201.5	6.4	2019	22	AAF61065	P. putida KT2440-a
38	174	5.5	2778	18	AAV25138	H. pylori cell env
39	174	5.5	2778	20	AAV75802	H. pylori outer me
40	173.5	5.5	27373	24	ABU91257	Chlamydia pneumoni
41	173.5	5.5	273254	21	AAAC81914	Chlamydia pneumoni
42	173.5	5.5	1230025	20	AAV91990	Nucleotide sequenc
43	173	5.5	3012	19	AAV52012	Helicobacter polyp
44	172.5	5.4	2751	22	AAF25588	H. pylori HPS120 e
45	169.5	5.3	2550	21	AAZ61502	DNA encoding the C

ALIGNMENTS

RESULT 1
AAA48507
ID AAA48507 standard; DNA; 1830 BP.
XX
AC AAA48507;
XX
DT 03-JAN-2001 (first entry)
XX
DE Neisseria meningitidis BASB040 putative coding sequence #1.
XX
KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
XX meningitis; cancer; autoimmune disease; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..1830
FT /*tag= a

FT XX /product= "BASB040"
 PN WO200034480-A1.
 PD 15-JUN-2000.
 XX 02-DEC-1999; 99WO-EP09560.
 PF 07-DEC-1998; 98GB-0026886.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Ruelle J;
 PI WPI; 2000-423426/36.
 XX P-PSDB; AAY99623.
 DR Novel BASB040 polypeptides of *Neisseria meningitidis* useful for
 PT diagnostic, prophylactic and therapeutic purposes against microbial
 PT diseases comprise a specific amino acid sequence -
 XX
 XX Claim 14; Page 59; 98pp; English.
 PS
 XX The present sequence is a putative version of the *Neisseria*
 CC meningitidis strain ATCC13090 BASB040 coding sequence. The protein
 CC produced from this gene is similar to the D15 outer membrane protein of
 CC the bacterium. The gene, its protein, antibodies, antagonists and
 CC agonists can be used to diagnose and treat bacterial diseases such as
 CC those leading to upper respiratory tract infections, bacteraemia and
 CC meningitis. In addition, they can be used in vaccines for use against
 CC cancer and autoimmune diseases.
 XX
 SQ Sequence 1830 BP; 440 A; 603 C; 461 G; 326 T; 0 other;

Alignment Scores:
 Pred. No.: 1.93e-285 Length: 1830
 Score: 3173.00 Matches: 609
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-857-669-2 (1-609) x AAA48507 (1-1830)

Qy 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
 DB 1 ATGATGATCAAAACCGACCGCCCTCTCTCGCGGCTTTATTTTCTTCGCGCAGCATC 60
 Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
 DB 61 GCGCCTGCGCGGACCTTTCCGAAACAAAGCGCGGTTTCGCATTGTTCCAAACAAA 120
 Qy 41 SerProAspThrGluSerValLysLysLysProLysPheProValArgIleAspThrGln 60
 DB 121 AGCCCCGACACCGAATCAGTTAAATTAACCCCAAAATTCGCCGCTCATCGACACGAC 180
 Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGlu 80
 DB 181 GATAGTGAATCAAGATATGTCGACAGACACCTCGCGCTCATCGCAGCAGCAGAA 240
 Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
 DB 241 GAAGTATTGGACAAGGAACACACGGGTTCTCGCGCAAGAACGACCGGACAGCTTAAA 300
 Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
 DB 301 ACCATGCTCCCGACGAAAGGCTATTTCAGCAGCAAGATTCAGCCCTGACGGAAGAACGCG 360
 Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
 DB 361 GCTTATACGGTACACATCACACCGCGCGCCGACCAAAATCGCCACGCTCGGTGTGCC 420
 Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160

Db 421 ATCCTCGCGACATCCTTTAGACGGCAACCTCGCCGAATACTACCGCAACGCGCTGAA 480
 Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
 DB 481 AACTGGCAGCAGCGCGTAGGCAGTGTTCGATCAGGACAGTTCGGGAAAACAGCAAACT 540
 Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
 DB 541 TCGTCTCTCGCGCGGTAAACGGAAAGGCTACCGCTTCCCAAGCTCGCAACACCCGG 600
 Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
 DB 601 GCGCCCTCAACCCCGATACCGCACCGTTCGATTGAACGTCGTCGTGGACAGCGCGC 660
 Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
 DB 661 CCCATCGCTTCGCGGACTTTGAAATCACCGGCACACAGCGTTTACCCGCAACAAATCGTC 720
 Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
 DB 721 TCGGCGCTGCGCGCTTCCAAACCGCGCACGCCCTACGACCTCGACCTGCTGCTGACTTC 780
 Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
 DB 781 CAACAGGCACCTCGAAACAAACCGGCAATTATTCGCGCGCTCCGTACAAGCGGACTTCGAC 840
 Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
 DB 841 CGTCTCAAGGCGACCGCGTCCCGTCAAAAGTCAGCGTAAACCGAGGTCAACAGCCACAAG 900
 Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
 DB 901 CTCGAAACCGCATCCGCTCGATTTCGGAATACGCTTTGGCGGCAAAATCGCCTAGCAC 960
 Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
 DB 961 TATTACAACCTCTTCAACAAAGGCTATATCGGCTCGGTCTGGATATGACAAATATC 1020
 Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 DB 1021 GAAACCGCTTTCGCGCGCATCAGCGCGGCAACTATCGGGCAACTACTGAGACA 1080
 Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
 DB 1081 AGCAACGTTTCTTCAACCGCTTCGACCCACCAAACTCGAAACCGCGCTTCTCGCGC 1140
 Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
 DB 1141 GGCATCTGGTATGTGCGCGACCGCGGCGCATCGATCCAGGCTGGGGCGAGAGTTTCTC 1200
 Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
 DB 1201 GCAGAGCGCGGAAATCCCGCGCTCGGATATCGATTTCGGCAACAGCCAGCCAGCATG 1260
 Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
 DB 1261 CTGACCGCTCTTGGAAACGCCAGCTGCTCAACACGCTGCTGCATCCGGAACAGCCAT 1320
 Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
 DB 1321 TACCTCAGCGCAAAATCGGTACGACTTTGGGCACATTCTCTGCTCCACCGCGCTGATC 1380
 Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
 DB 1381 CGCACCTCTGCGCGTGCAGGTTATTTCTACGCGCCGCAAAACAAACCTCGGACGCTTC 1440
 Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
 DB 1441 ATCATACGCGGACAAAGCGGTTACACCGTTCGCCGCGCAATCGCGAGCTTCTTCCAGG 1500
 Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520

Db 1501 CTGATGTTCCGACGGCGCGCTCTCCGTGCGCGTTACGAACCTGACAGCATCGGA 1560
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 1561 CTTGCGGCGCGAAGATCGTCTCCCGAAGCGCCCTCTCTGGTGGCGAGCTCGAA 1620
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 1621 TACCAACTGCGGTTTACGCGCACCTTCCGGCGCGGTTCACAGATATGGCGGATGCC 1680
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 1681 GCCGCCAATTTCAACGATATGAAGCTGAACACACGCTTCGGGACTTCGGCGTGGTTC 1740
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 1741 AGCCCGCTTGGCGGCTTCTTCGACATCGCCTACGGGCACAGCATAGAAATCCGC 1800
Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 1801 TGGCACATCAGCTTGGGAACAGCTTC 1827
RESULT 2
AAA48508
ID AAA48508 standard; DNA: 1830 BP.
XX AC AAA48508;
XX
DT 03-JAN-2001 (first entry)
XX
DE Neisseria meningitidis BASB040 putative coding sequence #2.
XX
KW BASB040: bacterial disease; respiratory tract infection; bacteraemia;
KW meningitis; cancer; autoimmune disease; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..1830
FT /tag= a
FT /product= "BASB040"
XX
FN WO200034480-A1.
XX
PD 15-JUN-2000.
XX
PF 02-DEC-1999; 99WO-EP09560.
XX
PR 07-DEC-1998; 98GB-0026886.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI; 2000-423426/36.
DR P-PSDB; AAY99624.
XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence
XX
PS Claim 11; Page 60; 98pp; English.
XX
CC The present sequence is a putative version of the Neisseria
CC meningitidis strain ATCC13090 BASB040 coding sequence. The protein
CC produced from this gene is similar to the D15 outer membrane protein of
CC the bacterium. The gene, its protein, antibodies, antagonists and
CC agonists can be used to diagnose and treat bacterial diseases such as
CC those leading to upper respiratory tract infections, bacteraemia and
CC meningitis. In addition, they can be used in vaccines for use against
XX cancer and autoimmune diseases.
XX
SQ Sequence 1830 BP; 439 A; 604 C; 461 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 1 93e-285 Length: 1830
Score: 3173.00 Matches: 609
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-857-669-2 (1-609) x AAA48508 (1-1830)
Qy 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 1 ATGATGATCAAAACGACGCGCTCTCTCGCGCTTTATTTCTTTCTCCGACGCATAC 60
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 61 GCGGCTCGCGCGACCTTTCGAAACAAAGCGCGGTTTCGCATTTGTTCAAAACAAA 120
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 121 AGCCCCGACACCGAATCAGTTAAATTAAACCCAAATTCCTCCGTCGCGATCGACGCGAG 180
Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 181 GATAGTGAATCAAAAGATATGGTCAAGAACACACCTCGCGCTCATCACGACGACGAGAA 240
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 241 GAATATTGGACAAGGAACACAGCGGCTTCTCGCCGAAGAACGCCGACACGTTAAA 300
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 301 ACCATGCTCCGACCAAGGCTATTTCAGCAGCAAAAGTCAGCCTGACGGAAAAAGACGGA 360
Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 361 GCTTATACGGTACACATCACACCGCGCGCGCCGACCAAAATCGCAACGTCGGTGTGCC 420
Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 421 ATCCTCGCGGACATCCTTTTCAGACGGCAACCTCGCCGAATACTACCGCAACGCGCTGGA 480
Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 481 AACTGGCAGCAGCGGTAGGACGTGATTTCGATCAGGCACAGTTGGGAAAAACAGCAAACT 540
Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 541 TCCGTCTCTCGCGCGGTAAACGCGCAAGGCTACCGGCTTGCAGAGCTCGCAACACCGCG 600
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValSerGlyArg 220
Db 601 GCGGCGGTCAACCGCGATACCGCCACCGCTCGATTGTAACGTCGTCGTGGACGCGCGCG 660
Qy 221 ProfileAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 661 CCCATCGCTTTCGGCGACTTTGAAATCACCGGCACACAGCGTTACCCGGAACAATCGTC 720
Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 721 TCCGCGCTTGGCGCGCTTCCAAACCGGGCACCGCCCTACGACCTCGACCTGCTGCTGACTTC 780
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 781 CAACAGGCACTCGAACAACGCGCATTTATTCGCGCGCTCCGTACAAAGCGACTTCGAC 840
Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 841 GCTCTCCAGGCGCGCGCTCCCGCTCAAGTCAGCGTAACCGAGGTCAACGCGCACCAAG 900
Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320

```
Db 901 CTCGAACCGGCATCCGCTCGATTCGGAATACGGTTCGGCGGCAAAATCCCTACGAC 960
Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 961 TATTACAACCTCTTCAACAAAGGCTATATCGGCTCGGTCTGGGATATGACAAATAC 1020
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 1021 CAACACAGCTTGGCGCGGCATCAGCCAGCGCGCAACTATCGGGGCAACTACTTGGACA 1080
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 1081 AGCAACGTTTCTCAACACGTTTCGACCAACCCAAAACCTCGAAAACCGGCTTCCTCCGGC 1140
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 1141 GGCATCTGGTATGTCGCGCAGCGCGGCATCGATCCAGGCTGGGGCAGAGTTTCTC 1200
Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 1201 GCAGAAGCGCGGAANAATCCCGGCTCGGATATCGATTTGGGCAACAGCCACGCCACGATG 1260
Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
Db 1261 CTGACCGGCTCTTGGAAACGCGCAGCTGCTCAACAACGTCGTGCATCCCGAAACGCGCAT 1320
Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 1321 TACCTCGACGCAANAATCGGTACGACTTTGGGCACATTCCTGTCTCCACCGCGTGATC 1380
Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 1381 CGCACCTCTGCGCGTGCAGGTATTCTTCACGCGCGAACAACAAAACTCGGCACGTTTC 1440
Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 1441 ATCATACCGGCAACAGCGGTTTACACCGTTGCCGCGCAATGCGGAGCTTCCTTCAGGG 1500
Qy 501 LeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 1501 CTGATGTTCCGACGCGCGCGGTCTTCGCTGCGCGGTACGAACTCGACAGCATCGGA 1560
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 1561 CTGCGCGCGCGCAACGATCGCTCTCGCGCAACGCGCGCTCTCGTGGGCGAGCGTGGAA 1620
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 1621 TACCAACTGCGGTTTACGCGCACCCCTTTCCGCGCGGTGTTCCACGATATGGCGCATGCC 1680
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 1681 GCGGCAATTTCAACAGTATGAAGTGAACACGGTTTCGGGAGTGGCGTGGCTGGTTC 1740
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 1741 AGCCCGCTTGGCGCGCTTTTCTCTCCATCGCTACGGCGCACAGCGGATGAAGAAATCGGC 1800
Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 1801 TGGACATCAGCTTGGGAACACGCTTC 1827
RESULT 3
AAZ53624
ID AAZ53624 standard; DNA; 1848 BP.
XX AAZ53624;
AC AAZ53624;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 286 partial DNA sequence SEQ ID NO:1197.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
```

```
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
OS Neisseria meningitidis.
XX WO957280-A2.
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
DR WPI: 2000-062150/05.
XX P-PSDB: AAZ74862.
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX Claim 7; Page 668-669; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 1848 BP; 440 A; 614 C; 468 G; 326 T; 0 other;
```

```
Alignment Scores:
Pred. NO.: 1,42e-283 Length: 1848
Score: 3153.00 Matches: 605
Percent Similarity: 99.51% Conservative: 1
Best Local Similarity: 99.34% Mismatches: 3
Query Match: 99.37% Indels: 0
DB: 21 Gaps: 0
```

US-09-857-669-2 (1-609) x AAZ53624 (1-1848)

```
Qy 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 19 ATGATGATTAAACCGACCGCTCTCTCTCGCGGCTTTATTTCTTCCGCGACGATAC 78
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 79 GCGCTGCGCGGACCTTTCCGAAAACAAAGGCGGGGTTTCGCATTGCTTCAAAACAAA 138
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 139 AGCCCCGACACCGAATCAGTTAAATTAANACCAATTTCCCGCTCGCATCGACGCGAC 198
```


61 AspSerGluLeuLysAspMetValGluGluHisLeuProLeuGluGlnGlu 80
199 GATAGTCAAAATCAAGATATGGTGAAGAACACCTGCGCTCATCAGCAGCAGGAA 258
81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
259 GAAGTATTTGACAGGAAACAGAGCGGCTTCTCGCCGAGAAAGACCGGCAACGTTAAA 318
101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
319 ACAATGCTCCGACGAAAGGCTATTTCAGCAGCAAGTACGCTGAGCGGAAAGACGGA 378
121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
379 GCTTATACGCTACACATCACACCGGCGCGCACCAAAATCGCAACGTCGCGTGGCC 438
141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrThrArgAsnAlaLeuGlu 160
439 ATCTTCGCGGACATCTCTTCAGACGGCAACCTCGCCGAATACTACCGCAACGCGCTGAA 498
161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
499 AACTGGCAGCAGCGGTAGCAGTATTCGATCAGACAGTATGGGAAACACGCAAACT 558
181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
559 TCGCTCTCGCGCGGTAAACGCAAGCGCTACCGCTTGCCAAAGCTCGGCAACACCGG 618
201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
619 GCGCGCTCAACCGCATACCGCACCGCGGATTTGAACGTCGCTGGGACAGCGCGC 678
221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
679 CCCATCCCTTCGCGGACTTTGAATTTACCGCAGCAGCGGTACCCGCAACAACTCGTC 738
241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
739 TCGGCTTGGCGGCTTCCAAACCGGCGCACGCTTACGACCTCGACCTGCTCGACTTC 798
261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
799 CAACAGCGCTCGAACAACAAACGGGCATTTATTCGCGCGCTCCGCTACAAAGCGACTTCGAC 858
281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
859 CGCTCCAGCGGACCGCGTCCCGCTCAAGTACAGCTAACCGAGGTCAACAGCGCAAG 918
301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
919 CTCGAAACCGGCATCCGCTCGATTTCGGAATACGTTTGGCGCGCAAAATCGCTACGAC 978
321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
979 TATTACAACCTCTTCAACAAGGCTATATCGGTTTCGCTCGCTGGGATATGCAAAATAC 1038
341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
1039 GAACACCGCTTGGCGCGGATATCGCGGCGCAACCTATCGGCGCAACTACTGAGCA 1098
361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
1099 AGCAACCTTCTTACACCGCTTCGACCAACCAACCTCGAAGGCGGCTTCTCGCGC 1158
381 GlyIleThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
1159 GGCATCTGGTATGTCGCGACCGCGCGGATCGATCGCAGGCTGGGCGGCGACTTCTC 1218
401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
1219 GCAGAGCGCGGAAATTCGCGGCTCGGATATCGATTGGGCAACAGCCACGCGCAGATG 1278
421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440

1279 CTCACCGCTCTTTGGAAACGCCAGCTCTCAACAACGCTGCTGCATCCCGAAACGGCAT 1338
441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
1339 TACTTCGACGGCAAAATCGGTACGACTTTGGCGCATTTCTCTCTCCACGCGCTCATC 1398
461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
1399 CGACCTCTCCCGCGCAGGTTATTTCTTCAGCGCGCAAAACCAAACTCGGACGCTTC 1458
481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
1459 ATCATACGCGACAAGCGGTTACACGCTTTCGCGCGCAATGCAACGTTCTCTCAGG 1518
501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
1519 CTGATGTTCCGACGCGCGCGCTCTTCTCGCGCGGTTACGAACTCGACAGCATCGG 1578
521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
1579 CTTGCGCGCGCAAGGATCGGTCCTCGCGCAAGCGCGCTCTTGGTGGGCGCTCGAA 1638
541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
1639 TACCACTGCGCTTTACGCGCACCTTTTCCGGCGCGTGTTCACACGATATGGGCGACGC 1698
561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
1699 GCGGCCAATTTCAACGATATGAAGCTGAAACACGTTTCGGGACTGGGCGTGGCTGTT 1758
581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
1759 AGCGCGCTCGCGGCTTTCTTCGACATCGCTTACGGGCGACACGCAAGAAATCCGC 1818
601 TrpHisIleSerLeuGlyThrArgPhe 609
1819 TGGCACATCAGCTTGGGAACGCGCTTC 1845
RESULT 4
AAZ53623 standard; DNA; 1848 BP.
ID AAZ53623 standard; DNA; 1848 BP.
XX
AC AAZ53623;
XX
DT~ 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 286 partial DNA sequence SEQ ID NO:1195.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR P-PSDB: AAY74861.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 7; Page 667; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA254941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 1848 BP: 439 A; 613 C; 467 G; 329 T; 0 other;

Alignment Scores:

Pred. No.: 5,42e-282 Length: 1848
 Score: 3136.00 Matches: 601
 Percent Similarity: 99.18% Conservative: 3
 Best Local Similarity: 98.69% Mismatches: 5
 Query Match: 98.83% Indels: 0
 DB: 21 Gaps: 0

US-09-857-669-2 (1-609) x AA253623 (1-1848)

Qy 1 MetMetileLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
 Db 19 ATGATGATCAAAACCGACCGCCCTCTCTCGCGGCTTTATTTTCTTCGCGACGCATAC 78
 Qy 21 AlaProAlaAlaLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
 Db 79 GCGCCTCGCGCGACCTTCCGAAACAAAGCGCGGCTTCGCATTTGTTCAAAACAAA 138
 Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
 Db 139 AGCCCCGACACCGAATCAGTCAAAATTAACCCAAATTCGCCGCTCTCATCGACACGCAG 198
 Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln 80
 Db 199 GACAGTGAATCAAGATATGTCGAAGACACCTCGCGCTCATCGACGACGAGAA 258
 Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
 Db 259 GAAGTATTGGACAAGGAACACACGCGGCTTCCTCGCGCAAGCGCGGCAACGTTAAA 318
 Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
 Db 319 ACGATGCTCCCGACAAAGGCTATTTCAGCAGCAAGTTCAGCCTGACGCGAAAGACGGA 378
 Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
 Db 379 GCTTATACGGTACACATCACACCGCGCGCGCCGACCAAAATCGCCACGTCGGCGTCC 438
 Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
 Db 439 ATCTCTGGCGACATCTCTTTCAGACGGAACCTCGCGCAATACTACCGCAACGCGCTGAA 498
 Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
 Db 499 AACTGGGACGACCGCGGTAGGCGACGATTTTCATCGACGACAGTTGGGAAACAGCAAACT 558

Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
 Db 559 TCCGTCCTCGCGCGGTAAACGGCAAGCCTACCGCTTCCCAAGCTCGGCAATACGCAG 618
 Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
 Db 619 GCGCGCTCAACCCCGATACCGCACCGCGATTTGAACGTCGTCGTGGACAGCGCGCG 678
 Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGluIleVal 240
 Db 679 CCCATCGCCTTCGCGGACTTTGAAATCACCGGCACACAGCGTTTACCCGCAACAAATCGTC 738
 Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPhe 260
 Db 739 TCCGCGCTTCGCGGCTTCCAGCCGGTATGCGGTACGACCTCGACCTGCTCGACTTC 798
 Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
 Db 799 CAACAGCGCTCGAACAACAAACGGGCATTTATTCGCGCGCTCCGTACAAGCGGACTTCGAC 858
 Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
 Db 859 GCGCTCCAAGCGCACCGCGTCCCGTCAAAAGTCAGCGTAACCGAGGTCAACACGCCACAAA 918
 Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
 Db 919 CTCGAACCGGCATCCGCTCGATTCCGAATACGCTTTGGCGGCAAAATCGCCTACGAC 978
 Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
 Db 979 TATTACAACCTCTTCAACAAAGGCTATATCGTTCCGTCGTGGATATGGACAAATAC 1038
 Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 Db 1039 GAAACCCAGCTTGCCTCGCGCATCAGCCGCGCAACTATCGGCGCACTACTGTGACA 1098
 Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
 Db 1099 AGCAACGTTTCTTACACCGCTTCGACCCACCAACCTCGAANAACCGCCTTCTCCGCG 1158
 Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
 Db 1159 GGCCTCGGTATGTGCGCGACCGCGGCGATCGATCCAGCTGGGGGGGGAATTTCTC 1218
 Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
 Db 1219 GCAGAAGCGCGAAATATCCCGCGCTCGGCTGTCTGATTTGGCAACACGCCACCGCATG 1278
 Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
 Db 1279 CTGACCGCTCTTGGAAAGCGCAGCTGCTCAACACCGTGTGTCATCCGAAACGGCCAT 1338
 Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
 Db 1339 TACCTCGACGCAAAATCGGTACGACTTTGGGCACATTTCTGTCTCCACCGCGCTGATC 1398
 Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
 Db 1399 CGCACCTCTGCGCGGTGCGGCTTATTTCTACGCGCGCAACAAACAACTCGGCGCGTTC 1458
 Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
 Db 1459 ATCATACGCGACAAAGCGGTTACACCGTTCGCGCGCAATCGGACGTTCTCTTCAGG 1518
 Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
 Db 1519 CTGATGTTCGCGACGCGCGCGCTCTTCCGTGCGCGTTTACGAACTCGACAGATCGGA 1578
 Qy 521 LeuAlaGlyProAsnGlySerValIleProGluArgAlaLeuLeuValGlySerLeuGlu 540
 Db 1579 CTTGCCGCGCGCAACGATCGGCTCTGCGCGAACGCGCCCTCTCTGTGGGAGCGCTGGAA 1638

Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 1639 TACCAACTGCGGTTTACGCGCACCTTTCCGGCGCGGTGTCCACGATATGGCGATGCC 1698
Qy 561 AlaAlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 1699 GCGGCCAATTTCAACGATGAAGCTGAACACGGTTTCGGGACTGGCGGTGGCGTTC 1758
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 1759 AGCCCGCTGGCGGTTTCTTCGACATCGCTACGGCGCACACCGATAGAAATCCGC 1818
Qy 601 TrpHisLeuSerLeuGlyThrArgPhe 609
Db 1819 TGGCACAATCAGCTTGGGAACGCGCTTC 1845
RESULT 5
AA81473
ID AA81473 standard; DNA: 92934 BP.
XX
AC AA81473;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_21 SEQ ID NO:21.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI: 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisseria infections, for example, N.gonorrhoea .
XX
PS Claim 7; Page 471-498; 1760pp: English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisseria bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 92934 BP; 24051 A; 24249 C; 21902 G; 22730 T; 2 other;
Alignment Scores:
Pred. No.: 1.14e-279 Length: 92934
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 21 Gaps: 0
US-09-857-669-2 (1-609) x AA81473 (1-92934)
Qy 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 38268 ATGATGATCAACACGACCGCCCTCTCTCGCGCTTTATTTCTTCTCCGACGCGATAC 38327
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 38328 GCGCTCCGCGCGACCTTTCCGAAACAAAGCGCGGTTCGATGTTTCAAAACAAA 38387
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 38388 AGCCCGACACCGAATCAGTCAAAATTAACCCCAATTCCTCCCTCATCGACACGAG 38447
Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 38448 GACAGTGAATCAAGATATGGTGAAGAACACCTCGCGCTCATCAGCGACGACGAGAA 38507
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 38508 GAAGTATGGACAGGAACAGACGCGGTCTCTCCCGCAAGAGCGCGCAACGTTAA 38567
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 38568 ACGATGCTCGCGACGAAGGCTATTTCAGCAGCAAGTTCAGCTTCAGGAAAAAGACGGA 38627
Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 38628 GCTTATACGGTACATCACCACCGCGCCGACCAAAATCGCAACAGTCGCGCTGCC 38687
Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 38688 ATCTCGCGGACATCTTTTCAGACGCAACCTCCCGCAATACTACCGCAACGCGCTGAA 38747
Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 38748 AACTGGCAGCAGCGGTAGGACGCGATTTTCATCAGCACAGTTGGGAAAAACAGCAAACT 38807
Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 38808 TCCGTCTCTCGCGCGGTAAACGCGCAACCTTACCCGCTTGCACAGCTCGGCAATACGCG 38867
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 38868 GCGCGCGTCAACCCCGATACCGCCCGCGGATTTGAACGCTCGTCTGGAGACGCGCGC 38927
Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 38928 CCCATCGCTTCGCGGACTTTGAAATCACCAGCACACGCGCTTACCCGCAACAAATCGTC 38987
Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 38988 TCCGCGCTTCGCGGTTTCCAGCGCGGTATCGCGTACGACCTCGACCTCTCTCGACATC 39047
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280


```
Qy 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 141935 ATGATGATCAAAACGACCGCCCTGCTCCTCGCGCTTTATTTTCTTCCGACGCATAC 141994
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 141995 GCGCTCCCGCGGACCTTCGGAACAACAGCGCGGGTTTCGCAATTGTTCAAAAACAAA 142054
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 142055 AGCCCCGACACCGAATCAGTCNAATTAACCCCAAAATCCCGCTCCTCATCGACACGAC 142114
Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln 80
Db 142115 GACAGTCAAAATCAAAATATGTCGAAGAACACCTGCGCGCTCATCGACGACGAGAA 142174
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 142175 GAAGTATTGGACAAGGAACAGACGGGCTTCCTCGCCGAAGAAGCGCGGACAAGGTTAAA 142234
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 142235 ACGATGCTCCGACCAAGAGCTATTTCAGCAGCAAAAGTACAGCTGACGGAAAAAGACGGA 142294
Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 142295 GCTTATACGTTACATACACACGGCGCCGCGCACCAAAATCGCCAAACGTCGGGCTGCC 142354
Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 142355 ATCCTCGCGGACATCCTTTTCAGACGGCAACCTCGCCGAATACTACCGCAACGCGCTGGAA 142414
Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 142415 AACTGGCAGCAGCGGTAGGCGCGATTCGATCAGCAGCAGTTCGGGAAAAACAGCAAACT 142474
Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 142475 TCCGTCCTCGCGCGGTAAACGCGAAAGCTACCGCTTCGCAAGCTCGGCAATACGAC 142534
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 142535 GCGGCGCTGCAACCCGATACCGCCACCGCGGATTTGAACGTCGTCGGACAGCGCGCG 142594
Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 142595 CCCATCGCCTTCGCGGACTTTGAAATCACCGGCGACACAGCGTTACCCCGAACAAATCGTC 142654
Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 142655 TCCGGCCTTCGGCGCTTCCAGCCCGGTATGCGGTACGACCTCGACCTGCTCGACTTC 142714
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 142715 CAACAGCGCTTCGAACAAAAGCGGCAATTATTCGCGCGCGTCCGTACAAAGCGGACTCGAC 142774
Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 142775 CGCTCCAGCGGACCGCGTCCCCGTCAAGTCAGCGTAACCGAGGTCAACACGCCACAAA 142834
Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
Db 142835 CTCGAAACCGGCACTCCGCTCGAATTCGGAATACGTTTGGCGCGCAAAATCGCTACGAC 142894
Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 142895 TATTACACCTCTTCAACAAAGGCTATATCGGTTTCGGTTCGCTCGGATATGCAAAATAC 142954
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 142955 GAAACCGCTTCGCGCGGCGCATCAGCGCGCGCAACTACTCGGGCAACTACTGGACA 143014
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
```

```
Db 143015 AGCAACGCTTTCTTACAAACCGTTGCACCAACCAAACTCGAAAAACCGCGCTTCTCCGGC 143074
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 143075 GCGCTCTGGTATGTGCGCGACCGCGCGGCATGCATGCCAGCTGGGGGGGAATTTCTC 143134
Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 143135 GCAGAACGCGGAAAAATCCCGCGCTCGGCTGTCGATTTGGCAACACGACGCCACCATG 143194
Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
Db 143195 CTGACCGCCTCTTGGAAACGCCAGCTGCTCAACAACGTCGTGCTCATCCGGAACGCGCAT 143254
Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 143255 TACCTCCAGCGCAAAATCGGTACGACTTTGGGCACATTCCTGTCTCCACCGCGCTGATC 143314
Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 143315 CGACCTCTCGCGCTGCGAGGTTATTTCTTCAGCGCCGAAAAACCAAACTCGGCACGTT 143374
Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 143375 ATCATACGCGACAAGCGGTTACACCGTTGCCCGCGACAATGCCGACGTTCTTCAAGG 143434
Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 143435 CTGATGTTCCGCACGCGCGCGCTTTCGCTGCGCGGTTACGAACTCGACAGATCGGA 143494
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 143495 CTTGCCGCGCGAACAGATCGGTCTGCCGAAACGCGCCCTCTCTGTTGGGCGAGCTGGAA 143554
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 143555 TACCAACTGCGGTTTACGCGCACCTTTTCCGCGCGGTGTTCACAGATATGGCGGATGCC 143614
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 143615 GCCGCCAATTTCAACGTATGAGCTGAACACACGTTTCGGGACTGGGCGTGGCTGGTT 143674
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 143675 AGCCCGCTTCGCGCGTTCCTTCGACATCGCCTACGGGCACAGCGATAGAAAAATCCGC 143734
Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 143735 TGGCACATCAGCTTGGGAACGCGCTTC 143761
```

RESULT 7

AAA48509
ID AAA48509 standard; DNA; 1764 BP.

XX
AC AAA48509;

DT 03-JAN-2001 (first entry)

DE Neisseria meningitidis BASB040 putative coding sequence.

KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
KW meningitis; cancer; autoimmune disease; ds.

OS Neisseria meningitidis.

XX Key
FH CDS Location/Qualifiers
1..1764

FT
FT CDS /*tag= a
FT /product= "BASB040"
FT /partial

XX PN WO200034480-A1.

XX	15-JUN-2000.	Qy	183	LeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAla	202
PD					
XX	02-DEC-1999; 99WO-EP09560.	Db	481	CTCGCGCGGTAAACGCGCAAGAGCTACCGCTTGCCAAAGTCGGAATACGACGCGGCC	540
PF					
XX	07-DEC-1998; 98GB-0026886.	Qy	203	ValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIle	222
PR					
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	Db	541	GTCACCCCGATACCGCCCGCGGATTTGAACGTCGTCGTGGACAGCGCGCCCATC	600
PA					
XX	Ruelle J;	Qy	223	AlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGly	242
PI					
XX	WPI; 2000-423426/36.	Db	601	GCCTTCGGCGACTTTGAAATCACCGGCACACAGCGTTACCCCGAACAATGCTCTCCCGC	660
DR	P-FSDB; AAY93625.				
XX	Novel BASB040 polypeptides of Neisseria meningitidis useful for	Qy	243	LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGln	262
PT	diagnostic, prophylactic and therapeutic purposes against microbial				
PT	diseases comprise a specific amino acid sequence	Db	661	CTTCGCGCTTTCCAGCCGGTATGCCGTACGACCTCGACCTGCTGCTGACTTCCAACAG	720
PT					
XX	Claim 11; Page 61; 98pp; English.	Qy	263	AlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu	282
PS					
XX	The present sequence is a putative version of the Neisseria	Db	721	CGCTCGAACAAACGGGCAATTTCCGGCGCGTCCGTACAGCGGACTTCGACGCGCTC	780
CC	meningitidis strain H41/76 BASB040 coding sequence. The protein	Qy	283	GlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGlu	302
CC	produced from this gene is similar to the D15 outer membrane protein of				
CC	the bacterium. The gene, its protein, antibodies, antagonists and	Db	781	CAAGCGGACCGCGTCCCGCTCAAGTCAAGTCAGGTACCGAGGTCAAAACGCCACAACCTCGAA	840
CC	agonists can be used to diagnose and treat bacterial diseases such as				
CC	those leading to upper respiratory tract infections, bacteraemia and	Qy	303	ThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyr	322
CC	meningitis. In addition, they can be used in vaccines for use against				
CC	cancer and autoimmune diseases.	Db	841	ACCGCATCCGCTCGATTCGGAATACGTTTGGCGGCAAAATCGCCTACGACTATTATAC	900
XX					
XX	Sequence 1764 BP; 424 A; 581 C; 451 G; 308 T; 0 other;	Qy	323	AsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThr	342
SQ					
	Alignment Scores:	Db	901	AACCTCTTCAACAAGGCTATATCGGTTCCGTTCGGATATGGACAAATACGAAACC	960
	Pred. No. : 3.93e-271 Length: 1764	Qy	343	ThrLeuAlaLaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn	362
	Score: 3019.00 Matches: 579				
	Percent Similarity: 99.15% Conservative: 3	Db	961	ACGCTTCCCGCGGATCGCCAGCGCAACTATCGGGCACTACTGGGCAAGCAAC	1020
	Best Local Similarity: 98.64% Mismatches: 5				
	Query Match: 95.15% Indels: 0	Qy	363	ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIle	382
	DB: 21 Gaps: 0				
	US-09-857-669-2 (1-609) x AAA48509 (1-1764)	Db	1021	GTTTCTTCAACCGTTTCGACCAACCAACCTCGAAACCGCGCTTCTCGCGCGCGCTC	1080
		Qy	383	TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu	402
		Db	1081	TGGTATCTGCGCGACCGCGCGGATCGATCCAGGCTGGGGGCGGAATTTCTCGCAAA	1140
		Qy	403	GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr	422
		Db	1141	GGCCGAAATCCCGCGCTCGGCTGTCGATTTGGGCACACGCCACCCACCATGCTGACC	1200
		Qy	423	AlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeu	442
		Db	1201	GCCTCTTGGAAACGCCAGCTGCTCAACAACGTCGCTGCATCCGAAACGGCATTTACCTC	1260
		Qy	443	AspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThr	462
		Db	1261	GACGGCAAAATCGGTACGACTTTGGGCACATTTCTGTCTCCACCGGCTCATCCGACC	1320
		Qy	463	SerAlaArgAlaGlyTyrPheThrProGluAsnLysLysLeuGlyThrPheIleIle	482
		Db	1321	TCTGCCGTGCGAGGTATTTCTTCCGCCCGCAAAACCAAACTCGGCACCTTCATCAT	1380
		Qy	483	ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet	502
		Db	1381	CGCGACAAAGCGGTTACACCGTTGCCCGCACAAATCCGACGTTCTTCAGGGCTGATG	1440
		Qy	503	PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla	522
		Db	1441	TTCCGACGCGCGCGCTCTTCCGTGCGCGGTTACGAACTTCGACAGCATCGGACTTGCC	1500
		Qy	523	GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln	542
		Db	1501	GGCCGCAAGCGATCGGTCTCCCGCAAGCGCCCTCTCTGGTGGCGAGCCTGGAATACCA	1560
		Qy	543	LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAla	562

|||||
Db 1561 CTCCCGTTTACGGCGACCCCTTTCCGGCGGTGTTCCACGATATGGCGGATGCGCGGCC 1620
QY 563 AsnPhelLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
|||||
Db 1621 AATTTCAACGATGAAAGCTGAACACGGTTCGGGACTGGCGGTGCGTGTTCAGCCCG 1680
QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHis 602
|||||
Db 1681 CTTGGCGCGTTTCTTCGACATCGCTACGGGCGACAGCGATGAAGAAATCCGCTGGCAC 1740
QY 603 IleSerLeuGlyThrArgPhe 609
|||||
Db 1741 ATCAGCTTGGGAACACGCTTC 1761
RESULT 8
AAZ53622
ID AAZ53622 standard; DNA; 1569 BP.
AC AAZ53622;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 286 partial DNA sequence SEQ ID NO:1193.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99NO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
WPI; 2000-062150/05.
DR P-PSDB; AAY74860.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
PS Claim 7; Page 666; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.
XX
SQ Sequence 1569 BP; 388 A; 522 C; 393 G; 266 T; 0 other;
Alignment Scores:
Pred. No.: 3.86e-237 Length: 1569
Score: 2653.00 Matches: 512
Percent Similarity: 99.42% Conservative: 2
Best Local Similarity: 99.03% Mismatches: 2
Query Match: 83.61% Indels: 1
DB: 21 Gaps: 0
US-09-857-669-2 (1-609) x AAZ53622 (1-1569)
QY 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 19 ATGATGATCAACACGACCGCCCTGCTCTCGCGGTATTTATTTCTTTCGACGCGATAC 78
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 79 GCGCTGCGCGCGACCTTTCGAAACACAGCGCGGGTTTCGATTGTTCAAAAGCAA 138
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 139 AGCCCCGACACCGCAATCAGTCAAAATTAACCCCAAAATTCGCCGTCGCGATCGACGAC 198
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 199 GACAGTGAATCAAGATATGGTGAAGAACACCTCGCGCTCATCAGCGACGACGAGAA 258
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 259 GAGGTTTGGATAAGGAACACAGCGGATTCTTCCCGAAGAACACCGGCAACGTTAAA 318
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 319 ACAATGCTCGCGACCAAGGCTATTTTCAGCAGCAAGCTCAGCTCAGCGAAAGACGGA 378
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 379 GCTTATACGGTGCATCATCACCGGGCCCGCCACCAAAATCGCCACACTCGCGGCTGCC 438
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 439 ATCTCGCGGACATCTCTTCAGAGCGCAACTCTCGCGCAATACTACCGCAACGCGCTGGA 498
QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 499 AACTGGCAGCAGCGCGTAGGCAGCGATTTCGATCAGGACACAGTTGGGAAAACAGCAAACT 558
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 559 TCCGTCTCGCGCGGTAAACGCGCAAGGCTTACCCGCTTGCCAAAGCTCGGCAACACCCGG 618
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 619 GCGCGCTCAACCCCGATACCGCCCGGATTTGAACGTCGTCGTGGACAGCGCGCG 678
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 679 CCCATTGCTTCGGCGACTTTGAAATCACCGGCGCACAGCGGTTTACCCGGAACAACGCTC 738
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 739 TCCGGCTGCGCGCGCTTCCAAACCGGCGCGCCCTTACGACCTTCGACCTCGACTTC 798
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 799 CAACAGCGCGCTCGAACAAACAGGGCATTTATTCGCGCGGTCCTCGTACAAGCGACTTCGAC 858
QY 281 ArgLeu-GlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 859 CGCCTCCCAAGGGGACCGCGTCCCGCTCAAAAGTCAAGTCAACCGAGGTCAAAACGCCACAA 918


```
Qy 300 sLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAs 320
|||||
Db 919 ACTGAAACGGCATCCGCTCGATTGGAATACGGTTGGCGCGCAAAATCGCCTACGA 978
Qy 320 pTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTy 340
|||||
Db 979 CTATTACAACCTCTTCAACAAGGCTATATCGGCTCGCTCTGCGATATGGACAAATA 1038
Qy 340 rGluThrThrLeuAlaAGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpTh 360
|||||
Db 1039 CGAAACACAGCTTCCCGCGGCAATCAGCCAGCCGCGCAACTATCGGGCGCAACTACTGGAC 1098
Qy 360 rSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGI 380
|||||
Db 1099 AAGCAACGTTTCTTACAACGCTTCGACCAACCCCAAAACCTCGAAAACCGCGCTTCTCGG 1158
Qy 380 yGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLe 400
|||||
Db 1159 CGGCATCTGGTATGTGGCGCAGCCGCGGCGATCGATCGCAGGCTGGGGCGGAATTCT 1218
Qy 400 uAlaGluClVArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMe 420
|||||
Db 1219 CGCAGAAGCGCGGAATCCCGGCTCGGATGTCGATTGGGCAACAGCCACGCCACGAT 1278
Qy 420 tLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHI 440
|||||
Db 1279 GCTGACCGCTCTTGGAAACGCCAGCTGCTCAACAACGTCGTGACCCCGAAACCGCCA 1338
Qy 440 sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuII 460
|||||
Db 1339 TTACTCTGACGGCAAAATCGGGAGGACTTTGGGCACATTCCTGTCTCCACCGCGCTAAT 1398
Qy 460 eArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPh 480
|||||
Db 1399 CCGCACCTCTCGCGCGAGGTTATTTCTTACCGCCGCAAAACAAACTCGGCACGTT 1458
Qy 480 eIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGI 500
|||||
Db 1459 CATCATACGGGACAAAGCGGTTACACCGTTGACCGCGACAAATGCCGATGTCCCTCGGG 1518
Qy 500 yLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
|||||
Db 1519 GCTGATGTTCCGACGCGCGCGGCTCTTCGTCGCGGTTTACGAAGTT 1567

RESULT 9
AAF61044
ID AAF61044 standard; DNA: 1374 BP.
XX
AC AAF61044;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF06604.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKF2-) DKF2 DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
```

```
XX WPI; 2001-192469/20.
DR
XX
PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria.
XX
PS Claim 1a; Page 94; 158pp; German.
XX
CC This invention describes novel DNA sequences (I) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (I); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (I) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (I) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
CC carrying one or more (I). (I), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (I),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 1374 BP; 264 A; 428 C; 429 G; 251 T; 2 other;

Alignment Scores:
Pred. No.: 4.18e-26 Length: 1374
Score: 384.50 Matches: 115
Percent Similarity: 42.52% Conservative: 81
Best Local Similarity: 24.95% Mismatches: 224
Query Match: 12.12% Indels: 41
DB: 22 Gaps: 11

US-09-857-669-2 (1-609) x AAF61044 (1-1374)
Qy 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaVal 186
|||
Db 43 GGTGACAACTCAATCAGCGGCATTACKAGGATGYCAAGCGTTGTATCCAGAACGACGCG 102
Qy 187 ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
:::|||||:::
Db 103 TCGCGCTATGGCTTCTTCAGTGGCGGCTTCAGTAGCAGCGCCCTGGCGCTGCAGCCCGAA 162
Qy 207 ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAsp 226
|||:::
Db 163 GCCGGTGTGGCGATATCGAACTGCTTACCAGATGGCGCGCTTATCGCTGGCGGCC 222
Qy 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
|||
Db 223 GTCACCTTCGGTGGCGACACACCCCTGGACGAGGACCTGCTGCAGCGCATGGTGTCTTC 282
Qy 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
:::|||||:::
Db 283 AAGCCGGGTACCCCTTACGACTCGGAACCTGGTCCGACAGTGCAACAGCAGCTGCAATFCG 342
Qy 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArg 286
:::|||||:::
Db 343 AGCGGCTATTTCGAAGGCGTGGCGTGTGTGATGGCGGCCCTGCTGTGTGGCGGAAGA 402
Qy 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg 306
|||||
Db 403 GTCCCGGTGGATGTTTCATCTTGAAACCGGTAAACACGACGACCATGGGCTTGGCGTGGC 462
```



```
QY 307 LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn 326
DB 463 TTCTCGACCGACGTCGGCGCGCGAGCAAGCAACCTGGACCGCCACCTGGGTCAACCCA 522
QY 327 LysGlyTyrIleGlySerValValTyrAspMetAspLys----- 339
DB 523 CAAGGCCAC-----AGCTATGCTGGGAACCGAAGCTGTCGGCCCGCGCCAGACAGTC 576
QY 340 -----TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
DB 577 GGCTGTGGTATGACATTCCTCGACCGCGCTGACGACAAAGTTGGCTTTGGCCGC 636
QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArg 376
DB 637 GGCTACCAGAACGAGGCTT-----GCCGGCACCGACGCTCAGCAAGCTG 684
QY 377 AlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
DB 685 TTGACGGTGGCGCCGCGAGTGGCAGCAAGCTGCCAGTGGCTGGCAGGGGTGATTTCG 744
QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
DB 745 CTCAGGTACCAACGCAAGAAATATCCCTG-----GGTACGACCTCCGGTTTGACCAACCTG 801
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
DB 802 CTGATCGCGCGCTGACCTTCTCTCCCTCCGCGAGT-----GACAAACCTATCGATCCG 855
QY 437 GluAsnGlyHisTyrLeu-----AspGlyLysIleGlyThr 448
DB 856 CACAAGCGGTATCGCTGCGTTCGAGTTCGATGTTCAAGGTGCCAAGGAGGCTGGTGTCCGAC 915
QY 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
DB 916 ACC-----AACCTGTGTGACGCGACGATGTTCTCAAGGCTGACCACTCGGCGCAC 969
QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla----- 486
DB 970 -----AACCAACCGCTTCTCGGAGCGGTACAGTTTGGTGGCAGTGCAACCAAT 1017
QY 487 GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
DB 1018 GGCTTC-----AAGAACAACATTCGCGCGCTGCGTCTTCTTCGCGGGT 1062
QY 507 GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGly 526
DB 1063 GCGCACGAGTGTGCGGCTTACGACTACCAACACCTGTGCGCGAAGAACAGCAGCGGT 1122
QY 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGlyThrGlnLeuProPheThr 546
DB 1123 GACCGTATCGCGCGCGCTACTTGGTGGCAGGAGTGTCCAGTACCAAGTATTCGCTGACC 1182
QY 547 ArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArg 566
DB 1183 GAAAAATGGCGGTGCGGACGTTGCTGCAACGCAAGCAACGCTTCAACGACCTGGAGCTG 1242
QY 567 MetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe 586
DB 1243 CCCAGCGCTCAAGACCGGTGCGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1302
QY 587 SerPheAspIleAlaTyrGlyHisSerAspLys-----LysIleArgTrpHisIleSerLeu 605
DB 1303 CGCTGACCTGGCCAGGCGCTGGATGACGAAGGGGGCATTCGCTGCACCTTTCCATG 1362
QY 606 Gly 606
DB 1363 GGG 1365
RESULT 10
AAA50269
ID AAA50269 standard; DNA; 1737 BP.
XX
```

```
AC
XX AAA50269;
DT 07-NOV-2000 (first entry)
XX Haemophilus influenza strain Rd KW20 BASB067 gene.
DE BASB067 gene; outer membrane protein; vaccine; antibiotic;
KW antibacterial; screening; infection; diagnosis; therapy; ds.
XX Haemophilus influenzae.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..1737
FT sig_peptide /tag= a
FT mat_peptide /tag= b
FT 67..1734
FT /tag= c
PN WO200047737-A1.
XX
PD 17-AUG-2000.
XX
PF 04-FEB-2000; 2000WO-EP00887.
XX
PR 09-FEB-1999; 99GB-0002880.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J, Thonnard J;
DR WPI; 2000-515059/46.
DR P-PSDB; AAY95820.
XX
PT BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
PT used for diagnosing and treating H. influenzae infections -
XX
PS Claim 3; Page 77-78; 87pp; English.
XX
CC The present sequence is that of the coding region of the BASB067
CC gene of Haemophilus influenzae strain Rd KW20. It encodes a
CC 578-amino acid surface expressed protein (see AAY95820) that is
CC recognised by the immune system. The invention relates to
CC recombinant materials and methods for the production of BASB067
CC polypeptides and polynucleotides, for use especially in therapeutic
CC and prophylactic vaccines. It also relates to methods for using
CC such polypeptides and polynucleotides in the prevention and
CC treatment of microbial diseases, in diagnostic assays for detecting
CC diseases associated with microbial infections, and assays for
CC detecting expression or activity of BASB067 polypeptides or
CC polynucleotides. A polynucleotide having at least 85% identity
CC to the present sequence can be used in the recombinant production
CC of BASB067 immunogenic polypeptides in transformed host cells, and
CC in vaccine compositions.
XX
SQ Sequence 1737 BP; 517 A; 309 C; 380 G; 531 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 5,46e-25 Length: 1737
Score: 374.00 Matches: 153
Percent Similarity: 36.88% Conservative: 93
Best Local Similarity: 22.94% Mismatches: 263
Query Match: 11.79% Indels: 158
DB: 21 Gaps: 18
```

US-09-857-669-2 (1-609) x AAA50269 (1-1737)

```
QY 3 IleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyrAlaPro 22
DB 16 CTAAACCTCAGCAGCTTTTATTA---GCATTAGTGTCTTCCCGCATTTGCTGACAA 72
QY 23 AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheAsnLysSerPro 42
DB 16 CTAAACCTCAGCAGCTTTTATTA---GCATTAGTGTCTTCCCGCATTTGCTGACAA 72
```



```
Db 1276 AAAAGTCAAGCATCAGCGCGTGGTTCGTACTTATGACAGAAAATCATCGTATC----- 1329
Qy 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSer 499
Db 1330 ---GTTGCTCGTCTGAATCGGTATTTACATACAAAGATATTTGAAAAATTCGCGCT 1386
Qy 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 1387 ACATGCGCTTTCTTTGCTGGTGGCGATCGTAGTGTGCGGTACGGCGTATATAAAAAAATT 1446
Qy 520 GlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 1447 GCGCTAAATAAGATAAGAAATTTGTTGGCGGCTCACGCTTGTACCACTTCTTTTA 1506
Qy 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 559
Db 1507 GAAATCAATATCAAGTTTATCCAAATTTGTTGGCGGCACTTTTGCAGATAGTGATTA 1566
Qy 560 AlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrp 579
Db 1567 GCTGCGGATTAATACACAGCAAGAGCTGCGTTATGCGACAGCGGTTGGTGGCTTGG 1626
Qy 580 PheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAsp----- 596
Db 1627 GCATGCGCAGTGGTGGCGATTAAATTTTGATATGCCACACCATTCGTGTATAAAGATAAC 1686
Qy 597 ---LysLysIleArgTrpHisLeuSerLeuGlyThr 607
Db 1687 AGCAAAAATATTCATATTTTACATCGGACTTGGTACC 1722
```

RESULT 12

```
AAAL15155
ID AAA15155 standard; DNA; 2505 BP.
```

```
AC AAA15155;
```

```
DT 21-AUG-2000 (first entry)
```

```
DE DNA encoding outer membrane protein (omp) 85.
```

```
KW Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
KW meningococcal infection; protective immune response; vaccine; ss.
```

```
OS Neisseria gonorrhoea.
```

```
XX Key Location/Qualifiers
```

```
XX RBS /*tag= a
```

```
FT 5'UTR 1..59
```

```
FT /*tag= b
```

```
FT CDS 60..2438
```

```
FT /*tag= c
```

```
FT /product= "outer membrane protein 85"
```

```
FT sig_peptide 60..122
```

```
FT /*tag= d
```

```
FT 3'UTR 2439
```

```
FT /*tag= e
```

```
XX WO200023595-A1.
```

```
XX 27-APR-2000.
```

```
XX 22-OCT-1998; 98WO-US22352.
```

```
XX 22-OCT-1998; 98WO-US22352.
```

```
XX (UYMO-) UNIV MONTANA.
```

```
XX Judd RC, Manning SD;
```

```
XX WPI: 2000-339694/29.
```

```
XX P-PSDB: AAY84946.
```

```
XX New isolated outer membrane protein 85 of Neisseria gonorrhoea and N.
PT meningitidis useful for vaccine, therapeutic and diagnostic
PT compositions for gonococcal or meningococcal infections -
XX Claim 6; Fig 2; 98pp; English.
```

```
XX The present sequence encodes an outer membrane protein (omp) 85 of
```

```
CC Neisseria gonorrhoea. The omp polypeptides and polynucleotides are
CC useful in compositions for use in the prevention, treatment and diagnosis
CC of non-symptomatic gonococcal infection or meningococcal infection and
CC symptomatic disease. They are also useful for the detection of
CC hybridisation complexes. Antigens and antibodies specific omp proteins
CC also provide diagnostic, therapeutic and prophylactic compositions for
CC the treatment or prevention of the infections described above. The
CC antibodies are useful for inducing a protective immune response in
CC humans or animals with N. gonorrhoea, N. meningitidis, or other
CC Neisseria species. The proteins, antibodies and polynucleotide
CC sequences of the present invention may also be used in the screening
CC and development of chemical compounds such as drugs or vaccines.
```

```
XX SQ Sequence 2505 BP; 669 A; 787 C; 621 G; 428 T; 0 other;
```

Alignment Scores:

```
Pred. No.: 3.29e-14 Length: 2505
Score: 260.50 Matches: 145
Percent Similarity: 36.55% Conservative: 86
Best Local Similarity: 22.94% Mismatches: 245
Query Match: 8.21% Indels: 157
DB: 21 Gaps: 28
```

```
US-09-857-669-2 (1-609) x AAA15155 (1-2505)
```

```
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys-----ThrMet 102
```

```
Db 685 GAAGCGACCGGTTCACCGCCAGAAATTCGCCCAAGACATCGAAAGATAACCGACTTCT 744
```

```
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
```

```
Db 745 ACCAGAACAC-GGCTACTTCGATTCCTCGTATCCGATACCGACATCAACACCAACAA 803
```

```
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
```

```
Db 804 GACAAAACGAGCAGACCATCAAAATCACCGTCCACGAAGCGGACGTTCCGCTGGGGC 863
```

```
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyr 155
```

```
Db 864 AAAGTGTGATTT-----GAAGCGGACACCAACGAAGTCCCC 899
```

```
Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnGlnProValGly 167
```

```
Db 900 AAGCGCGAACTGGAAAAACTGCTGACCATGAAGCCCGCAATGGTAGCAA----- 950
```

```
Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
```

```
Db 951 -----CGCCAGCAGATGACCGCGCTTTGGTGAGATTTCAG 986
```

```
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
```

```
Db 987 AACCGCATGGCTCGCGAGGCTACGATACAGCGAAATTC---AGCGTACACCGCGTCGCG 1043
```

```
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAla 223
```

```
Db 1044 AACCGCGGAACCAAAACCGTCGATTCGTCTGCACATCGAACCGCGGAGAAAATCTTAC 1103
```

```
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
```

```
Db 1104 GTCAACGAAATCCACATCACCAGCAACAAACAAACCCCGCAGGAGTGTGCGCGCGCAA 1163
```

```
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
```

```
Db 1164 TTGCGCCAAATGGAATCGCGCCCTTACGACACCTCCAAAGCTGCACACGCTCCCAAGAGCGC 1223
```

Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1224 GTCAGCTTTGGGCTACTTCGAC-----AACGTACAGTTTGATGCGCTGCCGTTCGC 1277
Qy 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1278 GGTACGCCGACCAAGTCGATTGACATAGAGCTGACCGAAGGTTCCACCGGCTGCTC 1337
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
Db 1338 GACTTGACCGCGGCTGGGTTACAGATACCGGCTTGTCATCGCCGCGGTATCCGAG 1397
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGlu 341
Db 1398 GACAACTGTTGGTACGGGC-----NAGTCGGCGCCCTCGCGCGCTCGCGAAGCAA 1451
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpTrpSer 361
Db 1452 ACCACGCTCAACGCTCGCTGCTCG-----TTTACCGACCGCTACTTCACGGCA 1499
Qy 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1500 GACGGGTGACGCTGGGCTAGATATTTACGGAAAGCCCTTCGACCCGCGCAAGCATCG 1559
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
Db 1560 ACCAGCTCAACAAATATAAACCACCACCGCGCGCGCGTAAAGATGGGTATCCCC 1619
Qy 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1620 GTTACCGAATACGACCGCTCAATTTCCGGCTGGCGCGGAACACCTGACCGTCAACACC 1679
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 1680 TACAAACAGACCCCAACCGTATCGCGACTTTATCAAAACATACGCCAAACACGACGGC 1739
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1740 GCAGAC-----GGCAGCTTCAAAGGCTGTGTACAAAGCCACTGTGCGGTGG 1787
Qy 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1788 GGGCGCAACAGACGACGCGCTTATGGCGCGCGCGC---TACCTGACCGGC--- 1841
Qy 446 IledGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1842 -----GTAATGCCGAAATTCGCGCTGCGCGCGCAACCTGCGCGCGCAAAA 1874
Qy 466 AlaGlyTyrPheThrProGluAsnLys-----LysLeuGly 478
Db 1875 CTGCAATACTACTCCGCCACCCACACCAACCTGTTCTCCCTTAAGCAAAACCTTC 1934
Qy 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1935 ACGTGATGTGCGCGCGGAAGTTCGCGGATTCGCGCGGCTACGCGCAGAACCAAGAAATC 1994
Qy 498 ProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1995 CCCTTCTTTGAAACTTCTACGGCGCGCGCTGGTGGTTCGGTGGCGGCTACGAAAGCGGC 2054
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 2055 ACGTCGGCGCGAAGTGTATGACGATACGCGCAAAATACGCTACGCGCGCAAC--- 2111
Qy 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2112 -----AAAAAGCCACAGCTCTCCGCCGAGCTGCTTCCCGATGCGCGGTGGC 2159
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2160 AAGACGACGACCGCTCGCGCTGAGGCTGTTTCCGACGACGACGCGTGTGGGACGGC 2219
Qy 560 -----AlaAlaAlaAsnPheLysArgMet----- 567

Db 2220 AGAACCCTATACCGCGCGCGAAACGCGTAACACAAATCGGTTTACTCGGAAACGCGCAT 2279
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2280 AAATCCACCTTTTACCAACGAATTCGCTATTCGCGCGCGCGGCTTACTCGCTCTCG 2339
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2340 CCTTTGGCGCGGATGAATTCATCTACCGCTACCGCTGAAGAAACCGGAAGACGAA 2399
Qy 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2400 ATCCAACGCTTCCAATTCACGCTCGGCACGACGCTC 2435
RESULT 13
AAS07278
ID AAS07278 standard; DNA; 2379 BP.
XX
AC AAS07278;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein DNA.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2379
FT /tag= a
FT /product= "N. meningitidis serogroup A antigen"
FT sig_peptide
FT /tag= b
FT mat_peptide 64..2376
FT /tag= c
FT /product= "Mature N. meningitidis serogroup A antigen"
XX
PN W0200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX WPI; 2001-381289/40.
DR P-PSDB; AAU03958.
XX
PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisserial bacteria infection -
PS Claim 6; Fig 2; 92pp; English.
XX

CC The sequence represents a DNA encoding a Neisseria meningitidis serogroup
CC A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx,
CC causing meningitis and, occasionally, septicaemia in the absence of
CC meningitis. This antigenic protein is useful in the manufacture of a
CC medicament for treating or preventing infection due to Neisseria
CC bacteria, such as meningitis and septicaemia. It is also useful as a
CC diagnostic reagent for detecting the presence of Neisseria bacteria or
CC antibodies raised against Neisseria, and as a reagent for raising the
CC antibodies. The Neisserial nucleotide sequences can be expressed in a
CC variety of different expression systems, for example, mammalian cells,

CC baculoviruses, plants, bacteria and yeast.

XX Sequence 2379 BP; 636 A; 755 C; 596 G; 392 T; 0 other;

Alignment Scores:

Pred. No.: 7, 23e-14 Length: 2379
Score: 256.50 Matches: 144
Percent Similarity: 36.39% Conservative: 86
Best Local Similarity: 22.78% Mismatches: 246
Query Match: 8.08% Indels: 157
DB: 22 Gaps: 28

US-09-857-669-2 (1-609) x AAS07278 (1-2379)

```

Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys-----ThrMet 102
Db 626 GAAGCGACCGCTTCGACCGCCAGAAATTCGCCCAAGACATCGAAAAAGTACCCGACTTCT 585
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 686 ACCAGAACAC-GGCTACTTCGATTTCGTTATCCTCGATACCGACATCCAAACCAACGAA 744
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAAACCGCGACGACCATCAAAATCACCGTCCACGAAGCGGACGCTTCGCGTGGGC 804
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAGTGTGCGATT-----GNAAGCGACACCAACGAAAGTCCCC 840
Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnProValGly 167
Db 841 AAGCGCGAACTGGAAAAACTGCTGACCATGAAGCCGCGCAATGGTACGAA----- 891
Qy 168 SerAspPheAsnGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCCGCTTTTGGGTGAGATTTCAG 927
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 928 AACCGCATGGGCTCGCGAGGCTACGCATACAGCGAATC---ACGCTACACCGCTCGCG 984
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223
Db 985 AACCGCGGAACCAACCGCTCGATTTCGCTGCACATCGAACCGCGCGGAAAAATCTAC 1044
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1045 GTCACGAAATCCACATCACCGGCAACACAAACCCGCGAGTGTGTCGCGCGCGAA 1104
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAla 263
Db 1105 TTGCGCCAAATGGAATCCGCGCCTTACGACACCTCCAAGCTGCAACGCTCCAAAGAGCGC 1164
Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1165 GTCAGAGCTTTTGGGCTACTTCGAC-----AACGTACAGTTTGTATGCGCTCCGCTTGC 1218
Qy 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACGCGCCGCAANGTCGATTGAACATGAGCCTGACCGACCGCTCCACCGGCTCGCTC 1278
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321
Db 1279 GACTTGAGCGCGGCTGGGTTTCAGGATACCGGCTTGTGTCATGTCGCGCGGCTATCGCAG 1338
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
Db 1339 GACAAACCTGTTGCGTACGGGC-----AAGTCGCGCGCCCTGCGCGGCTCGCGAAGCAA 1392
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 1393 ACCAGGCTCAACGGCTCGCTGTGCG-----TTTACCGACCGCTACTTCCAGCGCA 1440

```

RESULT 14

AAH42129

ID AAH42129 standard; DNA; 2379 BP.

XX

```

Qy 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1441 GACGGGTTCAGCTGGGCTACGATATTTACGAAAGCCCTTCGACCCCGCAAGCATCG 1500
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
Db 1501 ACCAGCGTCAAAACAATATAAAACACCAACCGCGCGCGCGGTAAAGGATGGGTATCCCC 1560
Qy 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGAATACGACCGCGCTCAATTTTCGGCTGCGCGGCAACACCTGACCGTCAACACC 1620
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 1621 TACAACAAAGCACCACCAACGCTATGCCGACTTTATCAGGAATACGCCAAACACGACGGC 1680
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1681 GCAGAC-----GGCAGCTTCAAAGCGCTGTGTACAAAAGCAGCAGCTCGGCTGG 1728
Qy 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GGGCGCAACAAAGACCGACGCGTCATGGCCGACGCGCGC---TACCTGACCGCGC--- 1782
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAANTCCGGAATCGCCCTGCCCGGACGAGAA 1815
Qy 466 AlaGlyTyrPheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAATACTACTCCGCGCACCAACACCTGTTCTTCCCTTTAAGCAAAACCTTC 1875
Qy 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACGCTGATGCTCGCGCGGAGTTCGGCATTCGCGCGGCTGCGGTTCGGTCCGCGGTACGAAAGCGC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCTCTCTTGGAAAACTTCTACGGCGGGCGCTGCGGTTCGGTCCGCGGTACGAAAGCGC 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACGTCGCGCGCAAGTGTATGCAATACGCGCAAAATACAGTACGCGCGCAAC--- 2052
Qy 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCCACGCTCTCGCGCGAGTGTCTTCCCGATGCCCGGTGCG 2100
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACGACCGCTCCGCTCGCGCTGAGCTGTTTCCGCGACGCGCGCTGGGACGGC 2160
Qy 560 -----AlaAlaAlaAsnPheLysArgMet----- 567
Db 2161 AGAACCTATACCGCGCGCAAAACCGTAACAAACATCGGTTTACTCGGAAACGCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2221 AATTCACCTTTACCAACGAATTCGCTATTTCCGCGCGCGCGGTACTCGCTCTCG 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CCTTTGGTCCGATAAATTCAGTACGCTACCGCTACCCGCTGAAGAAAACCCGGAAGAGAA 2340
Qy 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCAACGCTTCCAATTTCCAGCTCGGACGACGTTTC 2376

```



```
Db 1691 GCAGAC-----GGCAGCTTCAAGGCGCTGCTGTACAAAGGACCGTCCGGCTGG 1728
Qy 426 LysArgGlnLeuLeuAsnAspValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GGGCGCAACAGACGACGCGCTCATGGCGGCGCGGCGGCTACCTGACCGCGC--- 1782
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAATGCCGAATCGCCTGCGCGGCGGCAAA 1815
Qy 466 AlaGlyTyrPhePheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAATACTACTCGGCACCAACCAACCACTGGTTCTTCCCTTAAGCAAAACCTTC 1875
Qy 479 ThrPheIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACGCTGATGCTCGCGGCGAAGTCCGCTACGGCGGCGGTACGGCAGACCAAAAGAAATC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCTTCTTTGAAACTTCTACGGCGGCGCTGCTGCTGCGGTGCGGCTACGAAAGCGGC 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACGCTCGCGCGCAAGTGTATGACGAATACGGCGCAAAATTCAGTACGCGGCGCAAC--- 2052
Qy 527 SerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCAACGCTCTCCCGCAGCTGCTCTTCCCGATCCCGGTCGCG 2100
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACGACGCGTCCGCTGAGCTGTTGCGCAGCGCAGCGAGGTGGGACGCGC 2160
Qy 560 -----AlaAlaAlaAspPheLysArgMet----- 567
Db 2161 AGAACCTATACCGCGCGCGAAGCGGTAAACAACAAATTCGTTTACTCGGAAACGCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSer 581
Db 2221 AATCCACCTTTACCAACGAATTCCGCTATTTCCCGCGCGCGGCTTACCTGCTCTCG 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CCTTTGGGTCGGATGAATTCAGCTACGCTACCGCTGAAAGAAACCGGAAGACGAA 2340
Qy 598 LysIleArgTyrHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCCAACGCTTCCAATTCCAGCTCGGCACGACGTTTC 2376
```

```
RESULT 15
AAS07279
ID AAS07279 standard; DNA; 2394 BP.
```

```
XX AAS07279;
```

```
DT 23-OCT-2001 (first entry)
```

```
DE Neisseria gonorrhoeae antigenic protein DNA.
```

```
XX Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; ds; Neisseria meningitidis.
```

```
XX Neisseria gonorrhoeae.
```

```
XX Key Location/Qualifiers
FH CDS 1..2394
FT /*tag= a
FT /product= "N. gonorrhoeae antigen"
FT sig_peptide 1..63
FT /*tag= b
FT mat_peptide 64..2391
```

```
FT /*tag= c
FT /product= "Mature N. gonorrhoeae antigen"
PN WO200138350-A2.
XX 31-MAY-2001.
XX 28-NOV-2000; 2000WO-IB01851.
XX 29-NOV-1999; 99GB-0028197.
XX 09-MAR-2000; 2000GB-0005698.
XX (CHIR-) CHIRON SPA.
XX (STAT-) STATENS INST FOLKEHELSE.
XX Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX WPT; 2001-381289/40.
XX P-PSDB; AAU03959.
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
XX gonorrhoeae, useful in the manufacture of a medicament for treating and
XX preventing Neisserial bacteria infection -
XX Claim 6; Fig 2; 92pp; English.
XX The sequence represents a DNA encoding a Neisseria gonorrhoeae 85 kDa
XX antigenic protein. N. gonorrhoeae is closely related to N. meningitidis,
XX which colonises the pharynx, causing meningitis and, occasionally,
XX septicaemia in the absence of meningitis. This antigenic protein is
XX useful in the manufacture of a medicament for treating and preventing
XX infection due to Neisseria bacteria, such as meningitis and septicaemia.
XX It is also useful as a diagnostic reagent for detecting the presence of
XX Neisseria bacteria or antibodies raised against Neisseria, and as a
XX reagent for raising the antibodies. The Neisserial nucleotide sequences
XX can be expressed in a variety of different expression systems, for
XX example, mammalian cells, baculoviruses, plants, bacteria and yeast.
XX Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 1.39e-13 Length: 2394
Score: 253.50 Matches: 151
Percent Similarity: 36.50% Conservative: 99
Best Local Similarity: 22.04% Mismatches: 276
Query Match: 7.99% Indels: 159
DB: 22 Gaps: 27
```

```
US-09-857-669-2 (1-609) x AAS07279 (1-2394)
```

```
Qy 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
Db 470 TCGCGCGCAACCGCGCTGCACATCGACATCAGCATTGACGGGCAAAATCCGCCAAATCA 529
Qy 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
Db 530 CCGACATCGAATTTGAAGGCAACCAAGTCTATTCCGACCGCAAACTGATGGCGCAGATGT 589
Qy 73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaG1 93
Db 590 CGCTGACGCGGCGGCGCATTTGGACATGGCTGACACCAAGCAACCAATTCACAGCAGAGA 649
Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe---- 109
Db 650 AATTTGCCCAAGACATCGAAAAAGTACCGACTTCTACCGACAAACCGGCTACTTCGATT 709
Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisI 126
Db 710 TCCGCATCTCTCGATACCGACATCCAAACCAACGAAAGCAACCAACGACGACCATCAAAA 769
Qy 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
Db 770 TCACCGTCCAGGAAGCGGCGGCTTTCGTTGGGGCAAAAGTCTCATC----- 816
```



```
QY 144 spileUserAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
Db 817 -----GAAGGCGACCAACCAAGTCCCAAGCCGAACTGCTGA 865
QY 161 -----AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpG 176
Db 866 CCATGAAGCCCGCAATGTACGAA-----C 892
QY 176 luAsnSerLysThrSerValLeuAlaValThrArgLys-----GlyTyrP 192
Db 893 GCACAGATAGTCCCGCTTTGGTGAGATTGAGAACCCATGCGGCTCGGACGGTACG 952
QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
Db 953 CATACACGGAATC---AGGTACAGCCGCTGCAACAGCCGCAACCAACCGTCGATT 1009
QY 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
Db 1010 TCGTCTCTGCATCGAACCGGCGGAAATCTACGTCAACGAAATCCACATCACCGCA 1069
QY 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
Db 1070 ACACAAACCCGCGAAGTGTGCGCGGCAATTCGCCCAATGGAATCCGCGCTT 1129
QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
Db 1130 ACACACCTCCAGCTCCACGCTCCAAAGAGCGCTCGAGCTTTGGGCTACTTCAC- 1188
QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
Db 1189 -----AACGTACAGTTGTATGCTCCGCTCGCGGCACACCGCACAAAGTCGATTGA 1243
QY 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
Db 1244 ACATGACGCTGACGACAGCTTCCACCGCTCGCTCGACTTGAGCGCGGCTGGGTACAGG 1303
QY 310 lutyGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuLeuAsnLysLysTyrI 330
Db 1304 ATACCGCGCTGGTCATCTCCGAGCGCTTCCCAAGACACCTGTCGTACGGGC----- 1359
QY 330 leGlySerValTrpAspMetAspLysTyrGluThrThrLeu----- 344
Db 1360 --AAGTCGCGCGCTCGCGGCTCCAGCAAGCAAAACACGCTCAACGCGCTCGGTGCTG 1417
QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
Db 1418 TTACCGACCGCTACTTCACGCGACGCGGCTGAGCGCTGAGCTGTTTACGGAAGAAG 1477
QY 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
Db 1478 CCTTCGACCGCGCAAGCATCGACAGCATCAACAAATATAAACCCACACCGCAGCGG 1537
QY 373 euGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
Db 1538 CAGGCATCCGCTGAGCGTGCTTTACCGAATACGACCGCTGAATTCGTTTGTGTG 1597
QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
Db 1598 CAGAACACCTGACCGTCAACACCTTACAAAGCGCCCAACACATATGCGGACTTTATCA 1657
QY 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
Db 1658 AGAATACGCAAAACCGACCGCACAGAC-----GGCAGGTTCAAAGGCTGGCTGT 1708
QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnG 439
Db 1709 ACAAGGTACCGTCCGCTGGGGGCGCAACAAACCGCAGCGCTTATGGCCGACGCGG 1768
QY 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
Db 1769 GC---TACCTGACGGC-----GTCAACGCGCGAAA 1795
```

```
QY 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
Db 1796 TCGCCCTGCCGCGCAGCAAACTGCAATACTACTCCGCCACCCACCAACCTGGTTCT 1855
QY 476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
Db 1856 TCCCTTTAAGCAAAACCTTCACGCTGATGCTCGCGCGGAAGTCGGCATTCGCGCGGCT 1915
QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerV 511
Db 1916 ACGCAGAACCAAGAATAATCCCTCTTTGAAAACTTCTAGCGCGCGGCTGGTTCCG 1975
QY 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Db 1976 TCGCGGATACGAAGGGCAGCGCTCGTCCGAAAGTGTATGACGAATACGGCGAAAAA 2035
QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
Db 2036 TCAGCTACGCGCGCAAC-----AAAAAGCCACGCTCTCCGCGGAGTGC 2080
QY 540 lutyGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Db 2081 TCTTCGCGATGCCGCGCGGAAAGACGCGCACCGTCCGCTGAGCGCTGTTCCCGACG 2140
QY 557 etGlyAsp----- 559
Db 2141 CAGCAGCGTGTGGGCGGCANAACTACGACGACACAGCAGTTCCGCGACCGCGGCA 2200
QY 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
Db 2201 GGGTTCAAAACATTTACGCGCGCGCAATACCCATAAATCCACCTTTACCAACGAATTG 2260
QY 570 yshHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 2261 GCTATTCCGCGCGCGGCTTACCTGGCTCTCGCTTTAGGCCCGCATGAATTCAGCT 2320
QY 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
Db 2321 ACGCTACCGCTGAAGAAAAACCGGAGACGAAATCCACGCTTCCAATTTCCAACCTCG 2380
QY 606 lyThrArgPhe 609
Db 2381 GCACGAGCTTC 2391
RESULT 16
AAH42130
ID AAH42130 standard; DNA; 2394 BP.
XX AAH42130:
XX AC
XX XX
XX 17-SEP-2001 (first entry)
XX DE
XX Nucleotide sequence of a Neisseria serogroup A protein.
XX Serogroup A protein; outer membrane protein; Neisserial infection;
XX KW
XX vaccine; ss.
XX OS
XX Neisseria meningitidis.
XX Key
XX Location/Qualifiers
XX CDS
XX 1..2394
XX /tag= a
XX /product= "Neisseria serogroup A protein"
XX sig_peptide
XX 1..63
XX /tag= b
XX mat_peptide
XX 64..2391
XX /tag= c
XX WO200152885-A1.
XX 26-JUL-2001.
XX 17-JAN-2001; 2001WO-IB00166.
XX PF
```

```
XX 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX (CHIR-) CHIRON SPA.
XX Pizza M, Rappuoli R, Giuliani M;
XX WPI; 2001-451895/48.
DR P-PSDB; AAB84746.
XX
XX Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against Neisserial bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component.
XX
XX Disclosure; Page 70-71; 83pp; English.
XX
XX The present sequence encodes a Neisseria serogroup A protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a Neisseria
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO99/57280,
CC WO99/36344, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
CC TbpA, TbpB, PilC, OpA, or Omp85. The composition is used for making
CC a medicament for treating or preventing infection due to Neisserial
CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
CC bacteria or of antibodies raised against Neisserial bacteria; and/or
CC a reagent which can raise antibodies against Neisserial bacteria. It may
XX also be used as a vaccine.
XX
XX Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1.39e-13 Length: 2394
Score: 253.50 Matches: 151
Percent Similarity: 36.50% Conservative: 99
Best Local Similarity: 22.04% Mismatches: 276
Query Match: 7.99% Indels: 159
DB: 22 Gaps: 27
US-09-857-669-2 (1-609) x AAH42130 (1-2394)
Qy 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
Dy 470 TCGCCCGCAACCGCTCGACATCACATCAGATTGACGAGGGCAATCCCGCAATCA 529
Qy 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
Dy 530 CCGACATCGAATTGAAGGCACCAAGTCTATTCCGACCGCAAACTGATGCGGAGATGT 589
Qy 73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGl 93
Dy 590 CGCTGACGAAGGGGCAATTGGACATGGCTGACACGAAGCAACCAATTCAACGAGCAGA 649
Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe---- 109
Dy 650 AATTGGCCCAAGACATGAAAGTAACCGACTTCTACCAAGAACACCGGCTACTTCGATT 709
Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisI 126
Dy 710 TCCGCATCTCTGATAGACATCCAAACCAACCAACCAACCAACCAACCAACCAACCA 769
Qy 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGly 144
Dy 770 TCACCGTCCACGAAGGGAGCTTTCGCTTGGGGCAAGTCTCCATC----- 816
Qy 144 spIleLeuSerGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
Dy 817 -----GAAGGCGACACCAACGAAGTCCCAAGCGCAACTGGAACCACTGCTGA 865
Qy 161 -----AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSertrpG 176
866 CCATGAAGCCGCGCAAAATGGTACGAA-----C 892
176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
893 GCCAGCAGATGACCGCGTGGTGAGATTACAGACCGCATGGGTGCGCAGCGCTACG 952
192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
953 CATACAGGGAATC---AGCGTACAGCGGTGCCAAGCGCGCAACCAAAACCGTCGATT 1009
212 euAsnValValLysSerGlyArgProIleAlaPheGlyAspPheIleThrGlyT 232
1010 TCGCTCTGCATCAACCGCGCGGAAATCTAGTCAACGAAATCCACATCAACCGCA 1069
232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaAArgPheGlnProGlyThrProT 252
1070 ACAACAAACCCGCGCAGTAGTCTGCGCGCGCAATGGCGCAATGGAATCCCGCGCTT 1139
252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
1130 ACGACACTCCAAGCTGCAACGCTCCAAAGAGCGGTGCGAGCTTTGGCTACTTCGAC- 1188
272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
1189 ----AACGTACAGTTGATCGCTGCGCTTGGCGGCACACCGGACAAAGTCGATTGA 1243
290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyLleArgLeuAspSerG 310
1244 ACATGAGCCTCGACGAGCGTTCCACCGGCTGCTCGACTTGAGCGCGGCTGGGTACAGG 1303
310 luTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
1304 ATACCGGCTGGTCATCTCCGACGGCTTCCCAAGACCACTGTTCCGTACGGGC---- 1359
330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
1360 --AAGTCGGCGCGCTCGCGCTTCAGGAGCAAAACACGCTCAACGGCTGCGTGTGCT 1417
345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
1418 TTACCGACCGGTACTTCCGCGCAGCGGGTCAAGCTGGGTGCTGCTTTACCGAAAG 1477
358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
1478 CTTTCGACCCCGCAACGATCGACGACATCAACAATAATAAACCCACCGCAGCGG 1537
373 euGluLysArgAlaPheSerGlyGlyIleTyrThrValArgAspArgAlaGlyIleAsp 393
1538 CAGCATCCGTCATGAGGTGCTGTTACCGAATACGCGGTGAATTTTCGTTTGTGTG 1597
393 la-----ArgLeuGlyAlaGluPheLeu 401
1598 CAGAACCTCGTCGCTCAACACCTACAAACAAAGCGCCCAACATATGCCGACTTTATCA 1657
401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
1658 AGAAATACGGCAAAACCGCAGCGCACAGAC-----GGCAGCTTCAAGAGGTGGTGT 1708
421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnG 439
1709 ACAAGGTACGTCGCTGGGGGCGCAACAAACCGACGCGCTTATGGCGCAGCGCGG 1768
439 lyHisTyrLeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaL 459
1769 GC---TACCTGACGGGC-----GTGAACCGCCGAA 1795
459 euIleArgThrSerAlaArgAlaGlyTyrPheThrProGluAsnLys----- 475
1796 TCGCCCTGCGCGCAGCAAACTGCAATACTACTCCGCGCACCCACCAACCACTGGTTCT 1855
476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
```

```

Db 1856 TCCCTTAAGCAAAACCTTCACGCTGATGCTCGCGCGGCGAAGTCGCGATTCGGGGGGCT 1915
Qy 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerV 511
Db 1916 ACGGCAGAACCAAGAAATCCCTCTTTGAAACTTCTACGCGCGCGCTGGGTTCGG 1975
Qy 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Db 1976 TGGCGGATACGAAGACGCGACGCTCGTCCGAAAGCTGTATGACGAATACGGCGAAAAA 2035
Qy 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
Db 2036 TCAGCTACGGCGGCAAC-----AAAAAGCCCAACGTCTCCGCGAGCTGC 2080
Qy 540 LuTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Db 2081 TCTTCCCGATCCCGCGGCGAAGACGCGGACCGTCCCGCTGACCGCTGTTTGGCGAG 2140
Qy 557 etGlyAsp----- 559
Db 2141 CAGGCACGCTGTGGGACGCGCAAAACCTACGACGACACAGCAGTTCGCGGCGCGGCA 2200
Qy 560 -----AlaAlaAlaAsnPhelysArgHet-----LysLeuL 570
Db 2201 GGTTTCAAAACATTTACGCGCGCGCAATACCCATAAATCCACCTTTACCAACGAATTGC 2260
Qy 570 yHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 2261 GCTATTCGCGCGCGCGGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2320
Qy 590 leaAlaTyr-----GlyHisSerAspLysLysLysLysLysLysLysLysLysLys 606
Db 2321 ACGCCCTACCGCTGAAGAAAAACCGGAAGACGAGCAATCCCAACGCTTCCCAATCC 2380
Qy 606 lyThrArgPhe 609
Db 2381 GCACGACGCTC 2391

RESULT 17
AAS07277
ID AAS07277 standard; DNA; 2394 BP.
AC AAS07277;
XX AAS07277;
DT 23-OCT-2001 (first entry)
DE Neisseria meningitidis serogroup B antigenic protein DNA.
XX Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; ds.
XX Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2394
FT FT /tag= a
FT /product= "N. meningitidis serogroup B antigen"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..2391
FT /tag= c
FT /product= "Mature N. meningitidis serogroup B antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.

```

(STAT-) STATENS INST FOLKEHELSE.

Giuliani MM, Pizza M, Rappuoli R, Holst J;

WPI: 2001-381289/40.

P-PSDB; AAU03957.

Novel 85 kDa antigen from *Neisseria meningitidis* and *Neisseria gonorrhoeae*, useful in the manufacture of a medicament for treating and preventing *Neisseria* bacterial infection -

Claim 6: Fig 2; 92pp; English.

The sequence represents a DNA encoding a *Neisseria meningitidis* serogroup B 85 kDa antigenic protein. *Neisseria meningitidis* colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to *Neisseria* bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent for detecting the presence of *Neisseria* bacteria or antibodies raised against *Neisseria*, and as a reagent for raising the antibodies. The *Neisseria* nucleotide sequences can be expressed in a variety of different expression systems, for example, mammalian cells, baculoviruses, plants, bacteria and yeast.

Sequence 2394 BP; 657 A; 743 C; 583 G; 411 T; 0 other;

Alignment Scores:

Pred. No.: 1,54e-13 Length: 2394
Score: 253.00 Matches: 134
Percent Similarity: 36.25% Conservative: 90
Best Local Similarity: 21.68% Mismatches: 246
Query Match: 7.97% Indels: 148
DB: 22 Gaps: 25

US-09-857-669-2 (1-609) x AAS07277 (1-2394)

```

Qy 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPhe-----SerSer 111
Db 667 GAAAAGTAACTACGACTTCTACCAAAATACGGGTACTTCGATTCCGTATCCTCGATACC 726
Qy 112 LysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr-----ProGly 129
Db 727 GACATCCAAACCAACGAAGACAAACCAAGCAGACCATCAAAATCACCGTCCACGAAGC 786
Qy 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
Db 787 GCACGTTCGTTGGGCGCAAGTCTCCATC-----GAAAGC 822
Qy 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu-----Asn 161
Db 823 GACACCAACGAGTCCCAAGCCGAACTGGAATACTGCTGACCATGAAGCCCGGCAAA 882
Qy 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
Db 883 TGCTACGAA-----CGCCAGCAGATGACCGCC 909
Qy 182 ValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLysLeuGly 197
Db 910 GTTTTGGGTGAGATTGAGAACCGCATGGGCTCGCGAGGCTACGCATACAGCGAAATC--- 966
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
Db 967 AGCGTACAGCGCTGCGGAACGCTGAAACCAACCAACCGCTGCTGCTGCTGCTGCTGCTG 1026
Qy 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
Db 1027 CCGGCGCGGAAATCTACGTCAACGAAATACACATACCGCAGCAACAAACCCGCGAC 1086
Qy 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
Db 1087 GAAGTCGTCGCGGTGAATTTACGCCAAATGGAATCCGACCTTACGACACCTCAAGCTG 1146

```

QY 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
Db 1147 CAACGTTCCAAAGCGCGTTCGAGCTTTTGGGCTACTTCGAC-----AATGTCCAGTTT 1200
QY 278 AspPheAspArgLeuGlnGly-----AspArgValProValLysValSerValThrGlu 295
Db 1201 GATGCTGTCCGCGTTCGCGCAGCCGCGACAAAGTGCATTTGAACATGAGTCTGACCGAA 1260
QY 296 ValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
Db 1261 CGTTCACCGGTTCCCTGGATTGACGGCGGTGGTTCAGATACCGGGTTGGTCATG 1320
QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
Db 1321 TCCGCGAGGCTTCCCAAGACAACCTGTTCGGTACGGC-----AAGTCGGCGCGACTG 1374
QY 336 AspMetAspLysTyrGluThrThrLeu----- 344
Db 1375 CGCGCTCCAGAGCAAAACACGCTTAAGCGGCTCGTGTCTTACTGACCGGTACTTTC 1434
QY 345 ---AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp----- 359
Db 1435 ACGCGACAGGGGTACGCTGGCTAGATGTTTACGGAAGACCTTCGACCGCGCGAAA 1494
QY 360 ---ThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPhe 378
Db 1495 GCATCGACAGCATCAACAATATAAACACACCGAGCGCGAGCGCATCGCGATGAGC 1554
QY 379 SerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1555 GTCCCTGTTACCGAATACGACCGGTGAATTTCCGTTTGGTGGCAGAACACCTGACCGTC 1614
QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 1615 AACACCTACACAAAGCGCCCAACACTATGCGGACTTTATCAAGAAATACGCCAAAC 1674
QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSer 424
Db 1675 GACGGCACAGAC-----GCAGCTTCAAGGCTGGCTGTACAAAGTACCGTCGCGC 1725
QY 425 TrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGly 444
Db 1726 TGGGGCGGCACAAACACCGCGGCTTATGGCCGACGCGCGC---TACCTGACGGGC 1782
QY 445 LysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla 464
Db 1783 -----GTCAAGCGCGAATCGCCCTGCGCTGGCAGC 1812
QY 465 ArgAlaGlyTyrPheThrProGluAsnLys-----LysLeu 477
Db 1813 AACTGCAATACTACTCCGCGCACCCACCAACCAACCTGTTCTTCCCTCGAGCAAAAC 1872
QY 478 GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---Asp 496
Db 1873 TTCACGCTGATGCTCGCGCGGAAGTCGCGCATTTGCGGGCGGCTACGGCAGAACCAAGA 1932
QY 497 ValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeu 516
Db 1933 ATCCCTCTTTTGAANAATCTCTACGGCGCGGCTGGTTCGGTTCGGCGGATACGAAGC 1992
QY 517 AspSerIleGly-----LeuAlaGlyProAsn 525
Db 1993 GGCAGCTCGGTCGGAAGTCTATGACAAATACGGCGAATAATCAGTACGCGCGCAAC 2052
QY 526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro--- 544
Db 2053 -----AAAAAGCCCAACGCTCTCCGCGCAGCTGCTCTTCCCGATGCGCGCGC 2097
QY 545 -----PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2098 GCGAAGACGCGCGACCGTCCGCTGAGCCTGTTTGGCGCAGCGACGCGGTGCGGAC 2157
QY 559 ----- 559

Db 2158 GCGAAACCTTACGACGACAAACAGCAGTTCCGCGACCGCGGCTTCAAACATTTTAC 2217
QY 560 AlaAlaAlaAsnPheLysArgMet-----LysLeuLysHisGlySerGlyLeu 575
Db 2218 GCGCGCGGCAATACCCATAATCCACCTTTACCAACCAATTCGCTATTCGCGCGCGC 2277
QY 576 GlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
Db 2278 GCGGTTACCTGCGCTCGCGCTTTAGGCCCATGAAATTCAGCTACGCTACCGCTGGA 2337
QY 593 ---GlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2338 AAAAAACCGGACGAGCAATCCAAACGCTTCCAATTCCAACTCGCGCAGCAGCTTC 2391
RESULT 18
ID AAH42128 standard; DNA; 2394 BP.
XX
AC AAH42128;
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a *Neisseria* serogroup B protein.
XX Serogroup B protein; outer membrane protein; *Neisseria* infection;
KW vaccine; ss.
XX
OS *Neisseria meningitidis*.
XX
FH Key Location/Qualifiers
CDS 1..2394
FT /tag= a
FT /product= "Neisseria serogroup B protein"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..2391
FT /tag= c
XX
PN WO200152885-A1.
XX
XX 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-1B00166.
XX
PR 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX
XX (CHIR-) CHIRON SPA.
XX
PI Pizza M, Rappuoli R, Giuliani M;
XX
XX WPI; 2001-451895/48.
DR P-PSDB; AAB84744.
XX
PT Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against *Neisseria* bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component.
XX
PS Disclosure: Page 58-59; 83pp; English.
XX
XX The present sequence encodes a *Neisseria* serogroup B protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a *Neisseria*
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO95/57280,
CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
CC a medicament for treating or preventing infection due to *Neisseria*
CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*
CC bacteria or of antibodies raised against *Neisseria* bacteria; and/or

DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_26 SEQ ID NO:26.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 XX W0200022430-A2.
 XX
 XX 20-APR-2000.
 XX
 XX 08-OCT-1999; 99WO-US23573.
 XX
 XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Maignani V, Galcotti C, Mora M, Ratti G, Scarselli M, Scariato V;
 PI Rappuoli R, Pizza M;
 XX
 XX WPI: 2000-318079/27.
 DR
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 XX used in the diagnosis and treatment of N. meningitidis infection and
 XX other Neisserial infections, for example, N.gonorrhoea -
 XX
 XX Claim 7: Page 532-547; 1760pp: English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 XX represent specifically claimed Neisseria meningitidis genomic DNA
 XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 XX sequences, which are all used in the exemplification of the present
 XX invention. The nucleic acid sequences, protein sequences, and antibodies
 XX against them, can be used in the manufacture of a composition. The
 XX composition can be used as a medicament (or in the manufacture of a
 XX medicament) for treating, preventing or diagnosing infection due to
 XX Neisserial bacteria. For example, some of the identified proteins could
 XX be components of vaccines against Meningococcus B; against all serotypes;
 XX and/or against all pathogenic Neisseriae. Identification of sequences
 XX from the bacterium will also facilitate production of biological probes,
 XX particularly organism-specific probes. Attempts to make efficacious
 XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
 XX Multivalent vaccines have also been tried but none have successfully
 XX overcome antigenic variability. The provision of further, complete
 XX sequences may provide an opportunity to identify secreted or surface
 XX exposed proteins that may be presumed targets for the immune system and
 XX which are not antigenically variable or at least more conserved than
 XX other more variable regions.
 XX
 XX Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;
 SQ
 Alignment Scores:
 Pred. No.: 1.04e-11 Length: 52253
 Score: 253.00 Matches: 134
 Percent Similarity: 36.25% Conservative: 90
 Best Local Similarity: 21.68% Mismatches: 246
 Query Match: 7.97% Indels: 148
 DB: 21 Gaps: 25
 US-09-857-669-2 (1-609) x AAA81478 (1-52253)
 QY 97 AspaSnValLysThrMetLeuArgSerLysGlyTyrPhe-----SerSer 111
 Db 47964 GAAAAAGTAAACCGACTTCTTACCAAAATACGGCTACTTCGATTTCCTGATACC 48023

QY 112 LysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr-----ProGly 129
 Db 48024 GACATCCAAACCAACGAAGACAAACCAACGACACCATCAAAATCACCGCTCCACGAAGGC 48083
 QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
 Db 48084 GGCGTTTCGGTTGGGGCAAAAGTCTCCATC-----GAAGGC 48119
 QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu-----Asn 161
 Db 48120 GACACCAACGAAGTCCCAAGCCGAACCTGGAAACATGCTGACCATGAAGCCGGGCAAA 48179
 QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
 Db 48180 TGGTACGAA-----CCCGACGAGATGACCGCC 48206
 QY 182 ValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLysLeuGly 197
 Db 48207 GTTTTGGGTGAGATTCAAGACCGCATGGCTCGGCAGGCTACGCATACAGCGAAATC--- 48263
 QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
 Db 48264 AGCGTACAGCGCTGCCGAACCGCTGAAACCAACCAACCTCGATTTCTGCTGCACATCGAA 48323
 QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 Db 48324 CCGGGCCGGAATCTACGTCAAGCAATACACATACCGGCAACACAAACCCGGCAGC 48383
 QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 257
 Db 48384 GAAGTCTCGCGCTGAATTACGGCAATGGAATCCGACCTTACGACACCTCCCAAGCTG 48443
 QY 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
 Db 48444 CAACGTTCCAAAGAGCGCGCTCGAGCTTTGGGTACTTCCGAC-----AATGTCCAGCTT 48497
 QY 278 AspPheAspArgLeuGlnGly-----AspArgValProValLysValSerValThrGlu 295
 Db 48498 GATGCTCTCCGCTTGGCGGCGCCGCGCACCAAGCTCGATTGAACATGACTCTCACCGAA 48557
 QY 296 ValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
 Db 48558 CGTTCCACCGGTTCCCTGGATTTCGAGCGCGGTTGGGTTCAAGATACCGGTTGGTTCATG 48617
 QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
 Db 48618 TCCGCGAGGCTTCCCAAGACCACTGTTCGTACGGGC-----AAGTCGGCGGCACGTG 48671
 QY 336 AspMetAspLysTyrGluThrThrLeu----- 344
 Db 48672 CGCGCTCCAGGAGCAAAACACCGCTTAACGGCTCGCTGCTGTTTACTACCGCTACTTC 48731
 QY 345 ---AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp----- 359
 Db 48732 ACGCAGACGGGTCAGCTCGGCTAGCATGTTTACGGAAGGCTTCGACGCGCGCAAA 48791
 QY 360 ---ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPhe 378
 Db 48792 GCATCGACCAAGCATCAAAACAATATATAAACCAACACCGGCGCGCATCCCGCATGAGC 48851
 QY 379 SerGlyGlyIleTrpValArgAspArgAlaGlyIleAspAla----- 393
 Db 48852 GTGCTGTTACCGAATACGACCGGTGANTTCGGTTTGGTGGCAGACACCTGACCGCTC 48911
 QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
 Db 48912 AACACCTACAACAAAGCGCCCAACACTATGCGGCTTATCAAGAAATACGGCAAAACC 48971
 QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSer 424
 Db 48972 GACGGCACAGAC-----GGCAGCTTCAAGAGCTGGCTGTACAAAGGTACCGTCGGC 49022

```
QY 425 TrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGly 444
Db 49023 TGGGGGCGCAACAAACGACAGCGCTTATGGCGGACGCGCGC---TACCTGACGGC 49079
QY 445 LysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla 464
Db 49080 -----GTGAACGCCGAAATCGCCCTGCGCTGGCAGC 49109
QY 465 ArgAlaGlyTyrPhePheThrProGluAsnLys-----LysLeu 477
Db 49110 AAATGCAATACTACTCGGCCACCCACACCAACCTGGTTCTTCCCTCGAGCAAAACC 49169
QY 478 GlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---Asp 496
Db 49170 TTCACGTGTGCTCGCGCGGAGTCCGGATTGCGGGCGGCTACGGCAGAACAAAGAA 49229
QY 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
Db 49230 ATCCCTCTCTTTGAAACTTCTACGGCGCGGCGCTGGTTCGGTGGCGGATACGAAAGC 49289
QY 517 AspSerIleGly-----LeuAlaGlyProAsn 525
Db 49290 GGCAGCTCGTCCGAAAGTCTATGACGAATACGGCGAAATAATCAGCTACGGCGGCAAC 49349
QY 526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro--- 544
Db 49350 -----AAAAAGCCAAAGTCTCCGCCGAGCTGCTCTCCCGATGCCCGGC 49394
QY 545 -----PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 49395 GCGAAGACGCGCGCAGCTGCTGAGCGCTGTTCCCGACGAGCAGCGCTGTGGGAC 49454
QY 559 ----- 559
Db 49455 GGCAAAACCTACGACGACACAGCAGTTCCCGACCGCGGCGGCGGCTTCAAAACATTAC 49514
QY 560 AlaAlaAlaAsnPheLysArgMet-----LysLeuTyrHisGlySerGlyLeu 575
Db 49515 GCGCGCGCAATACCCATAAATCCACCTTTACCAACGAATTCGCTATTCGCGCGCGGC 49574
QY 576 GlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
Db 49575 GCGGTTACCTGGCTCTCGCTTTAGGCGCGGATGAAATTCAGCTACGCTACCCCTGAAG 49634
QY 593 ---GlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 49635 AAAAAACGGNAGACGAAATCCACGCTTCCCAATTCCAACTCGGACGACGTTTC 49688
```

RESULT 20

AAF21544/c

ID AAF21544 standard; DNA: 349980 BP.

XX AC AAF21544;

XX DT 13-MAR-2001 (first entry)

XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

XX KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;

XX KW ds.

XX OS Neisseria meningitidis.

XX PN WO200066791-A1.

XX PD 09-NOV-2000.

XX PF 08-MAR-2000; 2000WO-US05928.

XX PR 30-APR-1999; 99US-0132068.

XX PR 08-OCT-1999; 99WO-US23573.

XX PR 28-FEB-2000; 2000GB-0004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;

PI Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

DR Neisseria meningitidis B full length genome sequence and open reading

XX frames are used to detect, treat and prevent Neisserial infections -

XX Claim 7; Appendix A: 692pp; English.

XX The present invention describes the full length genome of

XX Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607

XX to AAF21613 represent fragments of the NMB genomic sequence, as the

XX sequences were too long to go in a record on its own it was split into 8

XX sequences which overlap each other at the beginning and end of each

XX sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at

XX the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at

XX the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

XX Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to

XX AAF21606 represent PCR primers which are used in the exemplification of

XX the present invention. The NMB genome and fragments from it have

XX antibacterial activity, and can be used in vaccines and gene therapy.

XX Neisseria nucleic acids, proteins and/or antibodies which binds to the

XX proteins can be used in compositions for treating or preventing infection

XX due to Neisserial bacteria or as a diagnostic reagent for detecting the

XX presence of Neisserial bacteria or of antibodies raised to Neisserial

XX bacteria. Computers, computer memory, computer storage medium or computer

XX databases can be used in a search to identify open reading frames (ORFs)

XX or coding sequences within the NMB genome. The DNA sequences provide

XX further opportunities to find antigenic or immunogenic proteins which are

XX more effective in vaccines than the outer membrane proteins currently

XX used.

XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

Alignment Scores:

Pred. No.:	1.4e-10	Length:	349980
Score:	253.00	Matches:	134
Percent Similarity:	36.25%	Conservative:	90
Best Local Similarity:	21.68%	Mismatches:	246
Query Match:	7.97%	Indels:	148
DB:	21	Gaps:	25

US-09-857-669-2 (1-609) x AAF21544 (1-349980)

QY 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPhe-----SerSer 111

Db 180140 GAAAAGTACCGACTTCTACCAAAATACCGCTACTTCGATTCCTGATTCCTCGATACC 180081

QY 112 LysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr-----ProGly 129

Db 180080 GACATCCAAACCAACGAGACAAACCAACGACGACCATCAAAATCACCGTCCACGAGGC 180021

QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149

Db 180020 GGACGTTTCCGTTGGGCAAAATCTCCATC-----GAAGGC 179985

QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu-----Asn 161

Db 179984 GACACCAACGAAAGTCCCAAGGCGAAGCTGGAAACCTGTCACCATGAAGCCCGGCAAA 179925

QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181

Db 179924 TGGTACGAA-----CGCCACGAGTACCGCC 179898

QY 182 ValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLysLeuGly 197

Db 179897 GTTTTGGGTGAGATTTCAGAACCGCATGGGCTCGGCGAGGCTACGCATACAGCGAAATC--- 179841

```
Oy 198 AsnThrArgAlaIaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
   : : : : :
Db 179840 AGCGTACAGCGCGTCCGCAACGCTGAACACCAACACCGTCGATTCGTCCTCCACATCGAA 179781

Oy 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
   : : : : :
Db 179780 CCGGCGCGGAAATCTAGCTCAAGCAATACATACACCGGCAACACAAACCCGCGAC 179721

Oy 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
   : : : : :
Db 179720 GAACTCGTCCCGCTGAATTCAGGCAATGAATCCGCACCTTACGACACCTCCAAAGCTG 179661

Oy 258 LeuAspPheGlnAlaLeuGlnAlaAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
   : : : : :
Db 179660 CAACGTTCCAAAGCGCGTCGAGCTTTGGGCTACTTCGAC-----AATGTCCAGTTT 179607

Oy 278 AspPheAspArgLeuGlnGly-----AspArgValProValIysValSerValThrGlu 295
   : : : : :
Db 179606 GATGCTGTCCCGCTTGGCGGCACGCCGACAAAGTCGATTTGAACATGAGTCTGACCGAA 179547

Oy 296 ValIysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
   : : : : :
Db 179546 CGTTCCACCGGTTCCCTGGATTTGAGCGCGGGTGGGTTCAAGATACCGGTTGGTCATG 179487

Oy 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
   : : : : :
Db 179486 TCCGCGCGGCTTCCCAAGACAACCTGTTCGTACGGGC-----AAGTCGCGCGCACTG 179433

Oy 336 AspMetAspLysTyrGluThrLeu-----
   : : : : :
Db 179432 CGCGCTCCAGGAGCAACACACGCTTAAACGGCTCGGTGCTTACTGACCCGCTACTTC 179373

Oy 345 ---AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp----- 359
   : : : : :
Db 179372 ACGGCACAGCGGCTGAGCTGGGTACGATGTTTACGGAAGAAAGCTTCGACCCGCGCAAA 179313

Oy 360 ---ThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPhe 378
   : : : : :
Db 179312 GCATCGACACGATCAACATATAAAACACACCGCAGCGCAGGATCCGATGACG 179253

Oy 379 SerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
   : : : : :
Db 179252 GTGCTGTGTACCGAATACGACGCGGTGAATTTTCGTTTGGTGGGAGAACACCTGACCGTC 179193

Oy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
   : : : : :
Db 179192 AACACCTACAAAGCGCCCAACACTATCCCGACTTATCAAGAAATACGGCAAAACC 179133

Oy 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSer 424
   : : : : :
Db 179132 GACGGCACAGAC-----GGCAGCTTCAAGGCTGGCTGTACAAAGGTACCGTCGCG 179082

Oy 425 TrpLysArgGlnLeuLeuAsnAsnValIleHisProGluAsnGlyHisTyrLeuAspGly 444
   : : : : :
Db 179081 TGGGGGGCAACAAACCGACAGCGGCTTATGGCGCGCGCGC---TACCTGACGCGC 179025

Oy 445 LysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla 464
   : : : : :
Db 179024 -----GTGAACCGCGGAAATCCGCCCTCGCTCGCTGGGAGC 178995

Oy 465 ArgAlaGlyTyrPhePheThrProGluAsnLys-----LysLeu 477
   : : : : :
Db 178994 AAACGTGAATACTACTCCGCACCCACCAACCAACCTGGTTCTTCCCTCCCTGAGCAAAACC 178935

Oy 478 GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---Asp 496
   : : : : :
Db 178934 TTCACGTGTGCTCGCGCGGGAAGTCGGCATTCGGCGCGGCTACGGCAGCAACCAAGAA 178875

Oy 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
   : : : : :
Db 178874 ATCCCTCTTTTGAANAATCTTACGGCGCGCGCTGGGTTGCGTGGCGGATACGAAGC 178815
```

```
Oy 517 AspSerIleGly-----LeuAlaGlyProAsn 525
   : : : : :
Db 178814 GGCAGCTCGTCCGAAAGTCTATGACGAATACGGCGAAANAATCAGCTACGGCGGCAAC 178755

Oy 526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro--- 544
   : : : : :
Db 178754 -----AAAAAGCCAAACGCTCTCCGCGAGCTGCTCTTCCCGATGCCCGGC 178710

Oy 545 -----PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
   : : : : :
Db 178709 GCGAAAGACGCGCAGCCGCTGAGCCTGTTTTCGCGAGCAGGAGCGTGTGGGAC 178650

Oy 559 -----
   : : : : :
Db 178649 GCGAAACCTACGACGACACAGCAGTTCCGCGCGCGCGCGGTTCAAAACATTTAC 178590

Oy 560 AlaAlaAlaAsnPheLysArgMet-----LysLeuLysHisGlySerGlyLeu 575
   : : : : :
Db 178589 GCGCGCGCAATACCCATAAATCCACCTTTTACCAACGAATTCGCGTATTCGCGCGCGGC 178530

Oy 576 GlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
   : : : : :
Db 178529 GCGGTTACCTGGCTCTCGCCTTTAGCCCGCATGAAATTCAGCTACGCGCTACCGCTGAAG 178470

Oy 593 ---GlyHisSerAspLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
   : : : : :
Db 178469 AAAAAACCGGAAGAGGAATCCACGCTTCCCAATTCCAACTCGGCACGACGCTTC 178416

RESULT 21
AAA15156
ID AAA15156 standard; DNA; 2394 BP.
XX
AC AAA15156;
XX
DT 21-AUG-2000 (first entry)
XX
DE DNA encoding outer membrane protein (omp) 85.
XX
KW Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
KW meningococcal infection; protective immune response; vaccine; ss.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2394
FT /tag= a
FT /product= "outer membrane protein 85"
XX
PN WO200023595-A1.
XX
PD 27-APR-2000.
XX
PF 22-OCT-1998; 98WO-US22352.
XX
PR 22-OCT-1998; 98WO-US22352.
XX
PA (UYMO-) UNIV MONTANA.
XX
PI Judd RC, Manning SD;
XX
DR WPI; 2000-339694/29.
XX
DR P-PSDB; AAY84947.
XX
PT New isolated outer membrane protein 85 of Neisseria gonorrhoeae and N.
PT meningitidis useful for vaccine, therapeutic and diagnostic
PT compositions for gonococcal or meningococcal infections .
XX
PS Claim 46; Page 85-89; 98pp; English.
XX
CC The present sequence encodes an outer membrane protein (omp) 85 of
CC Neisseria meningitidis. The omp polypeptides and polynucleotides are
CC useful in compositions for use in the prevention, treatment and diagnosis
CC of non-symptomatic gonococcal infection or meningococcal infection and
```


CC symptomatic disease. They are also useful for the detection of
 CC hybridisation complexes. Antigens and antibodies specific omp proteins
 CC also provide diagnostic, therapeutic and prophylactic compositions for
 CC the treatment or prevention of the infections described above. The
 CC antibodies are useful for inducing a protective immune response in
 CC humans or animals with N. gonorrhoeae, N. meningitidis, or other
 CC Neisseria species. The proteins, antibodies and polynucleotide
 CC sequences of the present invention may also be used in the screening
 CC and development of chemical compounds such as drugs or vaccines.

XX
 SQ Sequence 2394 BP; 656 A; 752 C; 587 G; 399 T; 0 other;

Alignment Scores:

Pred. No.: 2,13e-13 Length: 2394
 Score: 251.50 Matches: 151
 Percent Similarity: 36.50% Conservative: 99
 Best Local Similarity: 22.04% Mismatches: 276
 Query Match: 7.93% Indels: 159
 DB: 21 Gaps: 27

us-09-857-669-2 (1-609) x AAA15156 (1-2394)

Qy 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
 Db 470 TCGCCCGCAACCGCGTCGACATCGACATCGACGATTGACGGGCAAAATCCGCCAAATCA 529
 Qy 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
 Db 530 CCGACATCGAATTGAAGGCAACCAAGTCTATTCCGACCGCAAACTGATGGCGGAGATGT 589
 Qy 73 oLeuIleThrGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaG1 93
 Db 590 CGCTGACCGAAGCGGCGATTGGACATCGCTGCACGACGACGACCAATTCACGAGCAGA 649
 Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe----- 109
 Db 650 AATTGGCCCAAGACGAAAGTAACCGGACTTCTACGACGAACGCGTACTTCGATT 709
 Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaThrValHisI 126
 Db 710 TCGGTATCGTCTCGATCCGACATCCAAACCAACGAGGACGACAAACCAACGACCATCAAA 769
 Qy 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
 Db 770 TCACCGTCCACGAAGCGGAGCTTCCCTGGGGCAAGTCTCCATC----- 816
 Qy 144 spIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu----- 160
 Db 817 -----GAAGGCACACCAACGAAAGTCCCGCAAGCGGAACTGGAAAAAATCGCTGA 865
 Qy 161 -----AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpG 176
 Db 866 CCATGAACCGCGCAATGGTACGAA-----C 892
 Qy 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
 Db 893 GCCAGCAGATGACCCCGCTTTGGGTGAGATTGAGAACCGGATGGCTCGCGCAGGCTACG 952
 Qy 192 roLeuAlaLysLeuGlyAsnThrArgAlaValAlaAsnProAspThrAlaThrValAspL 212
 Db 953 CATACAGCGAAATC---AGCGTACAGCGCGTTCGCAAGCGGCAAAACCAAAACCGTCGATT 1009
 Qy 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
 Db 1010 TCGTCTCGACATCGAACCGGCGGCAAAATCTACGTCAACGAAATCCACATCACCGGCA 1069
 Qy 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaPheGlnProGlyThrProT 252
 Db 1070 ACACAAAACCGCGACGAGTGTGCGCGGGAATTCGCGCAATGGAATCCGGCGCTT 1129
 Qy 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerg 272
 Db 1130 ACGACACCTCCAAAGCTCAACGCTCAAAAGAGCGGCTGAGCTTTTGGCTTACTTCGAC- 1188

Qy 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
 Db 1189 -----ACGTACATTTGATGCGCTCCGCTTCCGCGCACACCCGACAAAGTCATTGA 1243
 Qy 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerg 310
 Db 1244 ACATGAGCTGACCGAAGCTTCCACCGGCTCGCTCGACTTGAGCGCGGCTGGGTACAGG 1303
 Qy 310 luTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
 Db 1304 ATACCGGCTGGTTCATGTCGCGAGCGGTTTCCCAAGACAACCTGTCGTTACGGGCGC- 1359
 Qy 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
 Db 1360 --AAGTCGGCGCGCTCGCGGCTTCAGGAAGCAAAACACGCTCAACGGCTCGCTGTCGT 1417
 Qy 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
 Db 1418 TTACCGACCGCTACTTTCAGCGCAGCGGCTCAGCTGGGCTAGCATGTTTACGGAAGA 1477
 Qy 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
 Db 1478 CCTTCGACCGCGCAAGCATCGACGACATCAACAATAATAAACACCAACGCGCGCG 1537
 Qy 373 euGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
 Db 1538 CAGGCATCCGATGAGCGTGTCTTACCGAATAGCAGCGGTGTAATTCGGTTGGTGG 1597
 Qy 393 la-----ArgLeuGlyAlaGluPheLeuA 401
 Db 1598 CAGAACACCTGACCGTCAACACCTTACAACAAAGCGCCCAACACTATGCCGCTTTATCA 1657
 Qy 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
 Db 1658 AGAAATACGGCAAAACCGACGCGCACAGAC-----GGCAGCTTCAAGAGCTGGCTGT 1708
 Qy 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
 Db 1709 ACAAGGTACCGTGGTGGGGGCGCAACAAACCGCAGCGGCTTATCGCGCGGCGCG 1768
 Qy 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
 Db 1769 GC---TACCTCAGCGGC-----GTGAACGCGCGAAA 1795
 Qy 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
 Db 1796 TCGCCCTCGCGCGGACGAACTGCAATACTACTCCGCCACCCACCAACCAACCTGGTTCT 1855
 Qy 476 -----LysLeuGlyThrPheIleArgGlyGlnAlaGlyTyrThrValAlaA 492
 Db 1856 TCCCTTAAAGCAAAACCTTACGCTGATGCTCGGGGCGCAAGTGGCATTCGGGCGGCT 1915
 Qy 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSrv 511
 Db 1916 ACGCAGAACCAAGAAATCCCTCTTTTGAATACTTCTACCGCGCGCGCTGGTTCGG 1975
 Qy 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
 Db 1976 TCGCGGATAGCAAGCGGACGCGTTCGGTTCGGAAGTGTATGCAATACGCGCAAAAAA 2035
 Qy 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuG 540
 Db 2036 TCAGTACCGCGGCAAC-----AAAAACCAACCTCTCCGCGGAGCTGC 2080
 Qy 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
 Db 2081 TCTTCCCATGCTCGCGCGAAGACGCGCGACCGTCCGCTGAGCTGTTTGGCGAG 2140
 Qy 557 etGlyAsp----- 559
 Db 2141 CAGGACGCTGTGGACGCGCAAAACCTACGACGACAAACAGGATTCCTCCGCGACCGCGCA 2200


```
Db 1422 -----AACAAATCTCTGGGTACACTGGCAC----- 1399
Qy 165 ProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 1399 ----- 1399
Qy 185 AlaValThrArgLysGlyTyProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsn 204
Db 1399 ----- 1399
Qy 205 ProAspThrAlaThrValAspLeuAsnVal---ValValAspSerGlyArgProIleAla 223
Db 1398 ---GATACAGCCGGTCACCGAAATGTCATCTAATGTATACACCA----- 1354
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyProGluGlnIleValSerGlyLeu 243
Db 1353 -----CCAGAGGGCCTTATATATACGACC 1330
Qy 244 AlaArgPheGlnProGlyThrProTyArgAspLeuAspLeuLeuAspPheGlnGlnAla 263
Db 1329 CCTCAGCGTCACCA---ACGCCAACTAAC-----AATCA 1297
Qy 264 LeuGluGlnAsnGlyHisTySerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1296 AGCCCAACAACCTCCACTGGTGGCCGCA----- 1267
Qy 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
Db 1266 GGAGAC-----GTAACC-----CGAAACCACCAACCGGCT 1234
Qy 304 GlyIleArgLeuAspSerGluTyArgLeuGlyLysIleAlaTyArgTyTrpAsn 323
Db 1233 GGACAGGACCGAGAAAT----- 1216
Qy 324 LeuPheAsnLysGlyTyIleGlySerValValTrpAspMetAspLysTyArgGluThrThr 343
Db 1215 -----ATGAGCGCCATGACACTATG 1195
Qy 344 LeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyArgGlyAsnTyTrp 359
Db 1194 CTCGAAAGGGATGTTGACCGCTCCATACATGTTCCGGGCTTCCGAGAGAAC----- 1141
Qy 360 ThrSerAsnValSerTyArgAsnArgSerThrThrGlnAsnLeu-----Glu 374
Db 1140 -----CAGAATCTCGTAGATCCTGACTGTACT 1114
Qy 375 LysArgAlaPheSerGlyGlyIle-TrpTyValArgAspArgAlaGlyIleAspAlaAr 394
Db 1113 CAGAAAGGTTTCCCTGGCTGTGAAGCGTTCAGTGGCGACCGGCGTACCCCGCTGCCGA 1054
Qy 394 gLeu-----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAs 412
Db 1053 TTTGTAATCCGGAGGCAGTTGGTGGCA-----CGGCTCGCTTATCATCGACGCTGCT 1000
Qy 412 pLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAs 432
Db 999 GATC-----ACCTTCATCTGCAATTCAAAACCGCGCGGCTGGC-AAATCA 956
Qy 432 nValLeuHisProGluAsnGly----- 439
Db 955 TGTGCTCATTTGCCGAACGGCATTTGACGTTACTGCTTCAGAACCCCGACATCACAC 896
Qy 440 -----HisTyLeuAspG1 444
Db 895 GCAAGGCGGCTGGCAGCGTCCATTAACTCGCTGGAGTCTCGACCACTTTACTCAGG 836
Qy 444 yLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAl 464
Db 835 TGAATTACCAATACCAAGCATGCTTTTATCTCTGGGCTGATGATTAGCCGACGCGGTC 776
Qy 464 aArgAlaGlyTyTrpPhe----- 469
Db 775 TCGTGGTGGCTGTATGCCAACCTGGGGGACTCGCAACGGTACTCTATCGACTACTCCAA 716
```

```
Qy 470 -----PheThrProGluAsnLysLysLeuG1 478
Db 715 CACGGCTGGGTTTCAGATGTCGATTTCTCCGTTTTCACGGCGAGACGCTGTGATCCG 656
Qy 478 yThr-----PheIleIleArgGlyGlnAlaGlyTyThrValIleAr 492
Db 655 CACATGTACGATGCCATCGTTTGTGTACACGGCGACGCTGGCTGGATTGAAACCGG 596
Qy 492 gAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValAr 512
Db 595 TGATTTGCACAAAGTACCGCGGATCTCGGTTTCTCCCGGGGCGACCGACGATTTTCG 536
Qy 512 gGlyTyTrpGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluAr 532
Db 535 TGGCTACAATACAATCTATCGCTCCGAATACCCCAACGCTGACCTGAAGAGGGGCTC 476
Qy 532 gAlaLeuLeuValGlySerLeuGluTyArgGlnLeuProPheThrArgThrLeuSerGlyAl 552
Db 475 GAAGTTGATAACCGGATCGCTGGAATACCACTGACCAACGTCACCGGAAATGTTGAGGCGC 416
Qy 552 aValPheHisAspMetGlyAspAlaAlaAlaAsnPhelysArgMetLysLeuLysHisG1 572
Db 415 GGTGTTTGTGATAGTGGCGAAGCGGTAAAGCATATTCCCGCGACGCGACTTTTAAACCGG 356
Qy 572 ySerGlyLeuGlyValArgTrp-PheSerProLeuAlaProPheSerPheAspIleAlaAr 592
Db 355 TACCGGGTTCGCGTGCCTGGGAATGCGCGCTCGGCCCAATCAAACTCGATTTTGGCG 296
Qy 592 yRGlyHisSerAspLysLys-----IleArgTrpHisIleSerLeuGly 606
Db 295 TACCGGTGCGGATAAAGCAACGCGGTTTACAGTTTATACATCGTCTGGGG 243
RESULT 24
ABQ72980
ID ABQ72980 standard; DNA; 2388 BP.
AC ABQ72980;
AC AC
DT 20-SEP-2002 (first entry)
DE
KW Methyllococcus capsulatus surface protein-antigen D15 DNA SEQ ID NO:5.
KW MopF; D15; surface protein antigen; helper protein; vaccine;
KW epitope mapping; gene; ds.
XX
OS Methyllococcus capsulatus.
XX
PN WO200255549-A2.
PD 18-JUL-2002.
PF 14-JAN-2002; 2002WO-NO00018.
PR XX
PR 12-JAN-2001; 2001NO-0000238.
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
DR WPI; 2002-557813/59.
XX
XX New nucleotide molecule encoding a surface exposed protein and is
PT incorporated and expressed in Methyllococcus capsulatus, useful in
PT manufacturing vaccines that can be administered orally in animals, fish
PT or humans
XX
PS Claim 1; Page 34-35; 39pp: English.
XX
```

CC The present invention describes a nucleotide molecule (I) comprising a
CC sequence that codes for any of the 14 Methylococcus capsulatus
CC nucleotide sequences given in AB072976 to AB072989. Also described is
CC a method for producing a desired protein in a bacterial host cell,
CC comprising: (a) transforming a bacterial host cell with a recombinant
CC vector comprising a first nucleotide sequence selected from the
CC group of 14 Methylococcus capsulatus, and comprising a nucleotide
CC encoding the desired protein linked in frame to the first nucleotide
CC sequence; and (b) culturing the transformed host cell in a medium under
CC conditions that allow the expression of the protein. (I) can be used
CC in vaccine production. The nucleic acids or the proteins from the
CC present invention can be used in manufacturing vaccines that can be
CC administered in animals, fish or humans. The method is useful in
CC producing fusion proteins carrying foreign peptide sequences that can
CC also be useful in epitope mapping, screening of antibody libraries and
CC as immunogens. The present sequence represents the specifically claimed
CC Methylococcus capsulatus surface protein-antigen D15 nucleotide sequence,
CC from the present invention.
XX

Sequence 2388 BP; 535 A; 653 C; 686 G; 514 T; 0 other;

Alignment Scores:
Pred. No.: 8, 1e-12 Length: 2388
Score: 234.50 Matches: 151
Percent Similarity: 32.98% Conservative: 99
Best Local Similarity: 19.92% Mismatches: 259
Query Match: 7.39% Indels: 249
DB: 24 Gaps: 29

US-09-857-669-2 (1-609) x AB072980 (1-2388)

QY 20 TyrAlaProAlaAlaAsp-----LeuSerGluAsnLysAlaAlaGlyPhe----- 34
DB 229 TACTCGCTGTCGGGAGGCGGACCCCTCGATGAAGAAACGGTCGCCGAGGTATCCGG 288
QY 35 AlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhePro 54
DB 289 GCGCTGTTCAAG-----ACGGGGTTTTTCAAG-----GAC 318
QY 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
DB 319 GTCCGCTGGACGAGGACGAGTAAAGTTGATCATCATCTACGTCGAAGAGCGT---CCTTCG 375
QY 75 IleThrGlnGlnGln-----GluGluValLeuAspLys 85
DB 376 ATTTCCAGCGTAAGATCGACGCGAACCATGACATCGGTACGAGGATCTGTAAGCGG 435
QY 86 GluGlnThrGlyPheLeuAlaGlu-----GluAlaProAspAsnVal 99
DB 436 CTCAAGGGAATCGGCTGGCGGAGGGAAGGTATTTCACCGCGCAGATCTCGACAAGGTC 495
QY 100 LysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAsp 119
DB 496 GAACAGAACTGGCGGCCAGTATTACAGCGCGGCAAAATACAGCCTT----- 543
QY 120 GlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAla---AsnValGly 138
DB 544 ---AAATCATTTCCCAAGGTGACGGAACCTGCCCGGNAACCGGTGGCGGTGAATATCAAT 600
QY 139 ValAla-----IleLeuGlyAsp----- 144
DB 601 ATCGCGGAGGCGGCTGCTCGATAAAACAGATCAATATCATCGGAATAATGCAATTT 660
QY 145 -----IleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
DB 661 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 156 ArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrp 175
DB 721 ACCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
QY 176 GluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLys 195
DB -----GACCATGGCAGCGAGTATTTCGACCTATTTCGCTCCTCCTCCTCCTCCTCCT 1757

DB 772 -----CTGGCGCTCTACTATCTCGATCGCGGTACGTCAATATTCGAA 813
QY 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
DB 814 ATCAGATCCACGAGTATCGATACGCCCAACAAGAAATATATATATATATATATATAT 873
QY 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGly---ThrGlnArg 234
DB 874 GTTAAGGAGGCGGAGCTCTTCAAGGTGGAGCAGGTACGCTGAGCGCAAGACCATCGTG 933
QY 235 TyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeu 254
DB 934 CCACCCGAACAACCTGCTCCCG---CTCGTGGTATCGGCGCGAGGATATATTTTCCAGG 990
QY 255 AspLeuLeuLeuAspPheGlnGlnAla----- 263
DB 991 AACTGGCAGACAAACCCAGAAAGGCAATTTCGATCTGTTGGCGAGGAGGCTATATC 1050
QY 263 ----- 263
DB 1051 TTCGCCAACGTCAATATGTTGCCGACATCAATCAGGAGAAGACCGTAACATCAC 1110
QY 263 ----- 263
DB 1111 TTTTTCGTCGATCCGGCAAAACAGGTCTACGTGAGCGGAATCAATTTCCAGGGCAATAC 1170
QY 263 ----- 263
DB 1171 AAGACGCTGACGAAGTGTGTCGACGGGAAATCGCGAGATGAGGCGGCTGGGCGTCG 1230
QY 264 -----LeuGluGlnAsnGlyHisTyrSerGlyAla 273
DB 1231 ACCCGCAAGATCCAGGCTTCCAGACACGAGCTCGAGGCTCGGTATTTCCAGGATGTC 1290
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
DB 1291 AAGCTGGAGCGCGCGGTGCGGGAACGACGAGTATCAGATCGACGCTCAATACATCAGTGTG 1350
QY 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
DB 1351 ACCGAAAGTCTTCGGCAACCTGACGCGGGGTGAGGGTATTCCCAAGTTCCAGGGGATC 1410
QY 314 GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuLeuAsnLysGlyTyrIleGlySerVal 333
DB 1411 ATATTCAATCGCGGCTCACCAGGACAACTTTTCGGAGCGGCAAGCGTGTAGTTC 1470
QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro----- 351
DB 1471 AATTTCAACAACAGCCAGATCAACACCATCTATGCCTTGGGGTATTTCATCTCTACGCC 1530
QY 351 ----- 351
DB 1531 ACCCTGGATGGCATCAGCAGTGTGTTGACATCAGTATCGCGACACCAACACCGGCTAT 1590
QY 352 ArgAsnTyrArgGlyAsnTyrTrpThr----- 360
DB 1591 TCNAACTATGTTCCCACTACATACCAATGTGTCAGGTGGCGGCAACTGGGGGTG 1650
QY 361 -----SerAsnValSerTyrAsn-----Arg 367
DB 1651 CCCATCGGAGAGTTCGACAGTATTTCGGACCAACTGAGTACAGTAAACACCAAGCTCAAG 1710
QY 368 SerThrThrGlnAsnLeuGluLys---ArgAlaPheSerGlyGlyIleThrTrpValArg 386
DB 1711 ACACCTCCAGCTTCCGATCAATCAATCAGCGCTTCAATGCTTCC----- 1752
QY 387 AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
DB 1752 ----- 1752
QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla---SerTrp 425
DB 1753 -----GACCATGGCAGCGAGTATTTCGACCTATTTCGCTCCTCCTCCTCCTCCTCCT 1757

Db 874 GTTAAGGAGCGGAGGCTTTCAAGGTGGACAGGTACGGCTACCGCGCAAGACCATCGTG 933
QY 235 TyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeu 254
Db 934 CCACCGAACAACCTGCTCCCG---CTCGTGGTATCGGGCGGAGGATATTTTTCAGG 990
QY 255 AspLeuLeuLeuAspPheGlnGlnAla----- 263
Db 991 AAACGTGGCAGACAAACCCAGAAAGGCGATTTCGATCGTTTGGCGAGGAGGCTATATC 1050
QY 263 ----- 263
Db 1051 TTCGCCAACGTCATATGTTGCCGACATCAATCAGSAGAAGACCGTAACATCAC 1110
QY 263 ----- 263
Db 1111 TTTTTCGTCTGATCGGGCAACAGGTCTACGTGAGCGCAATCAATTCAGGGCAATACC 1170
QY 263 ----- 263
Db 1171 AAGACGGCTGACGAAGTGTCTCGCAGGGAATGCGGCAGATGGAGGGCGCTGGGGCTGC 1230
QY 264 -----LeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 1231 ACCGCGAAGATCGAGCGTTCGAAGACCGCTCGAGGCTCGGCTATTTCCAGGATGTC 1290
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
Db 1291 AAGTGGAGAGCGCGGTGTCGGGAACGACGATCAGATCGACGTCAACTACAGTGTG 1350
QY 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
Db 1351 ACCGAGAAGTCTTCGGCAACCTGACGGCGGTGTAGGGTATTTCCCACTTCCAGGGCATC 1410
QY 314 GlyGlyLysIleAlaTyrAspTyrTyrAsnLeupheAsnLysGlyTyrIleGlySerVal 333
Db 1411 ATATTCAATCGCGGTCAACCCAGGACAACATTTTCGGCAGCGGCAACGCGTGTAGTTTC 1470
QY 334 ValTyrAspMetAspLysTyrGluThrLeuAlaAlaGlyIleSerGlnPro----- 351
Db 1471 AATTTCAACACAGCCAGATCAACACCATCTATGCTTGGGTATTTCAATCCTTACGCC 1530
QY 351 ----- 351
Db 1531 ACCCTGGATGCAATCAGCAGTGTTCGACATCAGTATCGCGACACCAACACCGGCTAT 1590
QY 352 ArgAsnTyrArgGlyAsnTyrTrpThr----- 360
Db 1591 TCGAACTATGTTGCCAACTACATTACCAATGTGTTCCAGGTGGCGGCAACTGGGGGTG 1650
QY 361 -----SerAsnValSerTyrAsn-----Arg 367
Db 1651 CCCATCGAGAGTTCGACAGTATTCGGACCAACCTGGACTACAGTAAACACCAAGCTCAAG 1710
QY 368 SerThrThrGlnAsnLeuGluLys---ArgAlaPheSerGlyGlyIleTyrTrpValArg 386
Db 1711 ACGACCTCCAGTCTTCGATCAATCAGACGCTTCAATGCTTCC----- 1752
QY 387 AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 1752 ----- 1752
QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla---SerTrp 425
Db 1753 -----GACCATGGCAGGAGTATTCGACCTATTCGTCCTCCTCACTGGGTGG 1797
QY 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGly-----HisTyrLeuAsp 443
Db 1798 ACCCATGACACCTGATCGGCGATATTCGGACACGCGGTGGCGCGCAGGATTCAGC 1857
QY 444 GlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 1858 GGACTGTTGCGCTGCGGTTTCAGCAGCCCTGACGTATTACAAGGGCAACGTCGCGCTGGAG 1917

QY 464 AlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg 483
Db 1918 CAG-----TATTTT-----CCGCTACCCAGATCTG---ACGCTGTGTTGAAT 1959
QY 484 GlyGlnAlaGlyTyr-----ThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 1960 GCGGATTTTCGTTACGGCGCGGCTATGGCAGTGGCGGCAACAGCGTCTTCGTTCTGG 2019
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 2020 GAACACTTTTACCGCGCGCTCCGAATTCGTCGCTGGATACACAGCCCAATTCACCTGGG 2079
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 2080 CCCAGGACACCGCTGCTACGCTTCGGCGGCAACAGCAAACTGACCGCTCGGTGGAG 2139
QY 541 -----TyrGlnLeuProPheThr-----ArgThrLeuSerGlyAlaValPhe 554
Db 2140 TTGCTGTTCGCTCCGTTCCGTTTCGGCGGAGAACTGAAGAGTGTCCGCTTGGGTACCTTC 2199
QY 555 HisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLys---LeuLysHisGlySer 573
Db 2200 GTCACGCGCGGAATGTTCGTCAAATTCGCCCAAGCTTTCGGACCTTCGCTTCGACC 2259
QY 574 GlyLeuGlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAla 591
Db 2260 GGCATTCGCCCAAGTGGCTATCACCTTCGGGGCCCTGATGTTCTCGATCGCC 2313
RESULT 26
AAS81897
ID AAS81897 standard; cDNA; 1425 BP.
XX AAS81897;
AC AAS81897;
XX 13-FEB-2002 (first entry)
DT
XX DNA encoding novel human diagnostic protein #17701.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG17710.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 17701; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques

DE		Genomic fragment #34.	
XX	KW	Genomic library; bacteria; human upper airway; otitis media; sinusitis;	
XX	KW	bronchopulmonary; endocarditis; meningitis; ss.	
XX	OS	Moraxella catarrhalis.	
XX	PN	WO200078968-A2.	
XX	PD	28-DEC-2000.	
XX	PF	16-JUN-2000; 2000HO-US15649.	
XX	PR	18-JUN-1999; 99US-0140121.	
XX	PA	(INCY-) INCYTE GENOMICS INC.	
XX	PI	Lagace RE, Patterson C, Berg KL;	
XX	DR	WPI; 2001-041427/05.	
XX	PT	Genomic library for identifying diagnostic and therapeutic	
XX	PT	compositions, and for identifying virulence factors, regulatory	
XX	PT	elements and drug targets, comprises Moraxella catarrhalis nucleic	
XX	PT	acids -	
XX	PS	Claim 1; Page 324-345; 545pp; English.	
XX	CC	The present invention relates to a Moraxella catarrhalis genomic library	
XX	CC	comprising of a combination of 41 nucleic acid molecules (see	
XX	CC	AAP28514-AAF2854).	
XX	CC	This library has a number of uses described in the	
XX	CC	specification e.g. is useful for identifying diagnostic and therapeutic	
XX	CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large	
XX	CC	aerobic, gram-negative diplococcus, normally found among the bacterial	
XX	CC	flora of human upper airways. M. catarrhalis is known to cause acute,	
XX	CC	localised infections such as otitis media, sinusitis and bronchopulmonary	
XX	CC	infection and life-threatening, systemic diseases including endocarditis	
XX	CC	and meningitis.	
XX	SQ	Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other:	
Alignment Scores:			
Pred. No.:	3.15e-08	Length:	89047
Score:	219.00	Matches:	130
Percent Similarity:	34.58%	Conservative:	74
Best Local Similarity:	22.03%	Mismatches:	232
Query Match:	6.90%	Indels:	154
DB:	22	Gaps:	18
US-09-857-669-2 (1-609) x AAP28547 (1-89047)			
QY	42	ProAspThrClusSerValLysLeuLysProLysPheProValArgIleAspThrGlnAsp	61
Dd	81780	CCTGTCAATGAACCGTTCAGTGTAACGGATGGATATTAAATGGATATTTGCCCATCGAA	81839
Qy	62	SerGluiLeLysAspMetValGluiHsiLeuProLeuLeuThrGlnGlnGluGlu	81
Dd	81840	TTAGTGCCATCTAATCTGATTCACACACAGCTAAATTTGGTGCTGCCAAGCCTGC CAT	81899
Qy	82	ValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAsnValLysThr	101
Dd	81900	TTATATGACATGCTCATGATAGGCTGCTTGCCATCAATCATGATGATGGCGTAAATCGC	81959
Qy	102	MeiLeuArgSerLysGlyTyrrPheSerSerLysValSerLeuThrGluLysAspGlyAla	121
Dd	81960	TCATT-----TTGGGCAGAATCACGGATGCCGATCTGCC-----	81995
Qy	122	TyrThrValHisilethrPROglyProArgThrLysilleAlaAsnValGlyValAlaIle	141
Dd	81996	-----GTTCACGTGCTATT	82010
Qy	142	LeuGlyAspLleLeuSerAspGlyAsnLeuAlaGluTyrtfArGrAsnAlaLeuGluAsn	161

QY 502 MetPheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAspSerIleGlyLeu 521
 DB 82824 CGTTTTTCTGGTGGCGACCAAGTATTCTGGATATCCACATGTTATCACCT 82883
 QY 522 AlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr 541
 DB 82884 ATATCAGATAGGCTTATCTGCACAGCGGTCACAGTATTGGCGGTGGTACAGCTGAATAT 82943
 QY 542 GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 561
 DB 82944 AATTATGAATTTATGAAGATTTGCGTTGGCGGTTTTTGGTGATATTGGTAAATGCTTAT 83003
 QY 562 AlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
 DB 83004 GATAAAGGCTTTACTAATGATACCAAAATTTGGTGCAGGTGTCGTCTGGGCATCA 83063
 QY 582 ProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys-----LysIle 599
 DB 83064 CCTGTCGGTCAAGTTCGTGTGTTGATGTGGCACTGGTCAAGAGAGGGCAATCCCATTT 83123
 QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
 DB 83124 AAGCTGCAATTTTATTGGCACACCATTT 83153
 RESULT 31
 ID AAQ66202
 XX AAQ66202 standard; DNA; 2989 BP.
 AC AAQ66202;
 XX 08-DEC-1994 (first entry)
 DT H. influenzae PAK 12085 D15 sequence.
 DE H. influenzae PAK 12085 D15 sequence.
 XX Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; NTHi; ss.
 XX Haemophilus influenzae PAK 12085.
 OS
 XX Key Location/Qualifiers
 FT CDS 389..2771
 FT /*tag= a
 XX
 XX W09412641-A.
 PN
 PD 09-JUN-1994.
 XX
 XX 23-NOV-1993; 93WO-CA00501.
 PF
 XX 23-NOV-1992; 92GB-0024584.
 PR
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX Chong P, Klein M, Loomore S, Sia DYC, Thomas W;
 PI Yang Y;
 PI WPI; 1994-200269/24.
 DR P-PSDB; AAR53758.
 DR
 XX Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 XX Disclosure; Fig. 1E; 161pp; English.
 PS
 XX Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca, PAK
 CC Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AAQ66198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale.

CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 2989 BP; 975 A; 468 C; 615 G; 931 T; 0 other;
 Alignment Scores:
 Pred. No.: 9.91e-10 Length: 2989
 Score: 213.50 Matches: 127
 Percent Similarity: 35.04% Conservative: 85
 Best Local Similarity: 20.99% Mismatches: 254
 Query Match: 6.73% Indels: 139
 DB: 15 Gaps: 24
 US-09-857-669-2 (1-609) x AAQ66202 (1-2989)
 QY 105 SerLysGlyTyrPheSerSerLysValSerLeuThr-----GluLys 118
 DB 1071 AATAATGGCTATGCCAAGCACAATCACTAAACAGGATGTTTACGTAAATGATGAAAAA 1130
 QY 119 AspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys----- 133
 DB 1131 ACAAAAGTTAATGTACCATTTGATGTAATCAAGGTTTACAGTATGACCTTCGTAGTGCA 1190
 QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGly 149
 DB 1191 CGCATTATAGGTAATCTGGAGGATGTCCTGCCGAGCTTGAACCTTTACTTTTACGACATTA 1250
 QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp 169
 DB 1251 CATTTAAATGATATCTTCCCGCCGTAGT----- 1277
 QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
 DB 1278 ---GATATTCAGATGTAGAAATGCAATTAAGCAAAACTTGGG-----GAACGA 1325
 QY 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp----- 206
 DB 1326 GGTTCAC-----GGTAACACACAGTAAATCTGTACCTGATTTTGGACGAT 1370
 QY 207 ---ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
 DB 1371 GCAAAATAAACATTTAGCGATAACCTTTGTTGTCGTCGACGACGCTTAACTGTTTCGC 1430
 QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
 DB 1431 CAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATAGTACTTTTCGTCAGGAATCGGA 1490
 QY 246 PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhePheGlnAlaLeuGlu 265
 DB 1491 CAACAAGAAGGAACCTTGGTATAATTCACAATAGTTAGGAAAAATTCGCTTAGAT 1550
 QY 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly--- 284
 DB 1551 CGTACAGGTTCTTCGAA-----ACAGTTGAAACCGAAATTTGATCTCATCAATGGTAGC 1604
 QY 285 ---AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
 DB 1605 AATGATCAAGTGGATGCTGCTATATAAAGTCAACAGCAGTACACGCGGTAGTATCACTTT 1664
 QY 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyrAspTyrTyrAsn 323
 DB 1665 GGTATTGTTACGGTACAGAGAGTGGTATCATAGTTATCAACAAGATTTAAACAAGATAAT 1724
 QY 324 LeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThr 343
 DB 1725 TTCTTTGGACAGCGGGCGAGTAGTAGTAAAGTATAGTAAATAATTTATGGTACGAGT 1784
 QY 344 LeuAlaAlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsn 357
 DB 1785 GTCAATTTGGGTTATACCGAACCCCTATTTTACTAAAGATGGTGAAGTCTTGGTGGAAAT 1844
 QY 358 TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg 376

```
Db 1845 ATTTCTTTGAAACTAGGATACTCTAAAGTGATACATCTCTAACTATAAAGCGTACG 1904
Qy 377 AlaPheSerGlyIle-----TrpTyrValArgAspArgAlaGlyIleAspAla 393
Db 1905 ACTTATGGAAGTAATGTTACTTTAGGTTTCCCTGTAATGAAATAACTCTTATATGTA 1964
Qy 394 ArgLeuGly-----AlaGluPheLeuAlaGlu----- 402
Db 1965 GGATTAGGCCATCTATATAAATAAGTAGTAACCTTTGCTGTAGATATAACCGTAATTAT 2024
Qy 403 -----GlyArgLysIleProGlySerAspIleAspLeuGly 414
Db 2025 TATATTCAATCAATCAAAATTTAAAGGTAAATGCGCATTAACAAATGACTTTGAT----- 2078
Qy 415 AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
Db 2079 -----TTTTCTTTGTTGGTGGAACTATACACAGCTTAATAGAGGCTAT 2120
Qy 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
Db 2121 TTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGTGGACGAGTTACTATTCCAGGT 2174
Qy 455 SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsn 474
Db 2175 TCTGATAACAATACTACAAGTAAAGTGCAGATGTACAGGGTTTCTACCCATTAGACAGA 2234
Qy 475 LysLysLeuGlyThrPheIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsn 494
Db 2235 GATCACCGCTGGTGTATCTGCAAAAGCATCTCGAGGATATGCAATGGTTTGGAAAC 2294
Qy 495 AlaAspValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyr 514
Db 2295 AAGCGTTTACCGTCTCTATCAAACTATATACAGCGGTGTCATTGTTCTATACCGGGTTT 2354
Qy 515 GluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2355 GCTTATGTGTATGTT-----GGGCTTAATGCAATTTATGCGCAACATGGTAATGGT 2405
Qy 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 2406 ACTTTTAATAAGATAAGTTCTGTATGTTGTTGTTGTAATGCAATCAACACTCGCGATCA 2465
Qy 540 GluTyrGlnLeuPro-----PheThrArgThr----- 548
Db 2466 GAACCTATTGTACCAACTCCATTTGTGAGTGATAAAAGCCAAATACAGTCCGAACCTCC 2525
Qy 549 -----Leu 549
Db 2526 CTATTTGTTGATGCGGCAAGTGTGTTGGAACTAATAATGGAAATCAGATAAAATGGATTA 2585
Qy 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLysLeu 569
Db 2586 GAGAGCAAGGCTCTCAAGAGACTTACCTGATTTATGCG-----AAATCAAGCGTATT 2636
Qy 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
Db 2637 CGCGCTCTACAGGTGTCGGATTCGAATGCGAATCTCCTATTGGACCATTTGTTCT 2696
Qy 590 IleAla-----TyrGlyHisSerAspLysLysIleArgTrpHisIleSer 604
Db 2697 TATGCTAAACCAANTTAAATAATATGAATATGATGTCGAA---CAGTTCCAAATTTAGT 2753
Qy 605 LeuGlyThrArgPhe 609
Db 2754 ATGGGGGCTCTTTC 2768
RESULT 32
ABK52155
ID ABK52155 standard; DNA; 944 BP.
XX
AC ABK52155;
XX
DT 13-AUG-2002 (first entry)
```

```
XX
DE
XX
KW BASB207: antibacterial; anti-inflammatory; vaccine; ss;
KW bacterial infection; influenza; pneumonia; bronchitis; sinusitis;
KW otitis media; infection; immunisation; antibacterial.
XX
OS Haemophilus influenzae.
XX
PN W0200232946-A2.
XX
XX 25-APR-2002.
XX
PF 16-OCT-2001; 2001WO-EPI1983.
XX
PR 17-OCT-2000; 2000GB-0025488.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Thonnard J;
XX
XX WPI; 2002-435533/46.
XX
XX New BASB207 polypeptides and polynucleotides, useful as vaccines
XX against bacterial infections particularly those caused by non-typeable
XX Haemophilus influenzae, e.g. pneumonia, sinusitis and otitis media -
XX
XX Disclosure; Page 82; 116pp; English.
XX
XX This invention relates to the cDNA and protein sequences of a novel
XX BASB207 protein. The proteins of the invention may have antibacterial or
XX anti-inflammatory activity and may be used to create a vaccine. The
XX BASB207 polypeptides and polynucleotides are useful as vaccines against
XX bacterial infections particularly those caused by non-typeable H.
XX influenzae, including pneumonia, exacerbation of chronic bronchitis,
XX sinusitis and otitis media; as immunogens to produce antibodies
XX immunospecific for such polypeptides and polynucleotides. These may also
XX be used in diagnosing the stage of infection and type of infection the
XX pathogen has attained. The polynucleotides may be further used in
XX genetic immunisation, as components of polynucleotide arrays useful for
XX diagnostic and prognostic purposes, in the discovery and development of
XX antibacterial compounds, and to interfere with the initial physical
XX interaction between a pathogen and a eukaryotic host responsible for
XX sequelae of infection. Compositions comprising the sequences of the
XX invention are useful in the preparation of a medicament for generating
XX an immune response in an animal. The present sequence represents the
XX upstream DNA sequence of the Haemophilus influenzae BASB207 cDNA of
XX the invention.
XX
SQ Sequence 944 BP; 274 A; 170 C; 220 G; 280 T; 0 other;
Alignment Scores:
Pred. No.: 2,54e-10 Length: 944
Score: 212.50 Matches: 58
Percent Similarity: 51.88% Conservative: 25
Best Local Similarity: 36.25% Mismatches: 69
Query Match: 6.70% Indels: 9
DB: 24 Gaps: 3
US-09-857-669-2 (1-609) x ABK52155 (1-944)
Qy 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr--- 471
Db 456 TGGCTATCAGAAATCTCTTTTATAAAGTCAAGCATCAAGCGGTGATTCGTAATAT 515
Qy 472 ProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAla 491
Db 516 GCAGAAAATCATCCTGATTT-----GTTGCTGCTGCTGAAATCGGGTATTTTACATACA 566
Qy 492 ArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerVal 511
Db 567 AAGATATTGAAAAAATTCGCCCTACACTGCGTTTCTTTGCTGGTGGCGATCGCAGTGTG 626
```

```
QY 512 ArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGlu 531
Db 627 CGTGGTTACGGCTATATAAAATTCGCCCTAAAAATGGAATGGGAAATGGTGGGTGC 686
QY 532 ArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGly 551
Db 687 TCACGTTTCTACAGGTTCTTTAGAAATATCAATATCAAGTTTATCCGAAATGGTGGTG 746
QY 552 AlaValPheHisAspMetGlyAspAlaAlaAlaAsnPhelysArgMetLysLeuLysHis 571
Db 747 GCACCTTTTCAGATAGTAGTAGCGCTAA-TATTACACGCAAAAGAGCTGCTTAT 805
QY 572 GlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla 591
Db 806 GCGCAGGCGTTGGTGTGGCATCCCGAGTGGCGCGATTAAATTTGATATTTGCC 865
QY 592 TyrGlyHisSerAsp-----LysLysIleArgTrpHisIleSerLeuGlyThr 607
Db 866 ACACCATTCGCGATAAGATAACACGCAAAATATTTCAATTTTACATTTGGACTTGGTACA 925
RESULT 33
ID AAQ66201 standard; DNA; 2974 BP.
AC AAQ66201;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. Influenzae SB33 D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; NTHi; ss.
XX
OS Haemophilus influenzae SB33.
XX
FH Key Location/Qualifiers
FT CDS 386..2764
FT CDS /*tag= a
XX
PN W09412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 92GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI; 1994-200269/24.
DR P-PSDB; AAR53757.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
PS Disclosure; Fig. 1D; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2974 BP; 971 A; 471 C; 615 G; 917 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 1.87e-09 Length: 2974
Score: 210.50 Matches: 121
Percent Similarity: 34.88% Conservative: 75
Best Local Similarity: 21.53% Mismatches: 241
Query Match: 6.63% Indels: 125
DB: 15 Gaps: 21

US-09-857-669-2 (1-609) x AAQ66201 (1-2974)
QY 134 IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu 151
Db 1193 ATAGGTAATCTCGGAGGTATGTCTGCCAGCTTGAACCTTTACTTTTACACATTACATTTA 1252
QY 152 AlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAsp 171
Db 1253 AATGATATCTTCGCGGTAGT-----GAT 1276
QY 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
Db 1277 ATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTTGGG-----GAACGAGGTTAC 1327
QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
Db 1328 -----GGTAACACACAGCATAAATTTCTGTACCTGATTTTGACGATGCAAAAT 1372
QY 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 1373 AAMACATTTAGCGATAACCTTTGTTGTGTGATGCTGGAGAGCGTTTAACTGTTCCACCACTT 1432
QY 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 1433 CGCTTTGAAGGAATAACCGTTTCTGCTGATAGTACTTTACGTCAGGAATGCGCCCAACA 1492
QY 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
Db 1493 GAAGGAACCTGGTATAATTCACAAATTAGTTAGTTAGGAATAATTCGCTTAGATCGTACA 1552
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
Db 1553 GGTTTCTTCGAA-----ACAGTTGAAACCGGAATGTATCTCATCATGTTAGTCAATGAT 1606
QY 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
Db 1607 GAAGTGGATGCTGATATATAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGTTATT 1666
QY 306 ArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
Db 1667 GGTACGGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTCAACAAGATAATTTCTTG 1726
QY 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAla 345
Db 1727 GGAACAGGGCGGCAGTAACTATAGTGGTACGAAAATGATTATGTTAGTACAGTGTCAAT 1786
QY 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTrp 359
Db 1787 TTGGTTATACCGAGCCCTATTTTACTAAAGATGGTGAAGTCTTGGTGGAAATGTTTTC 1846
QY 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArgAlaPhe 378
Db 1847 TTTGAAACACTACGATAACTCTAAAGTGATACATCTCTTAACATAAGCGTAGCTTAT 1906
QY 379 SerGlyClyIle-----TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
Db 1907 GGAAGTAATGTTACTTTAGTGTTCCTGTAAATGAAATAACTCTTATTTATGTAGGATTA 1966
QY 396 Gly-----AlaGluPheLeuAlaGlu-----402
Db 1967 GGCCATACCTATATAAAATTAGTAACCTTTCCTAGAAATAATACCTTAATTTATATT 2026
QY 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 2027 CAATCAATGAAATTTAAAGGTAATGGCATTAACAAACAAATGACTTTGAT-----2074
```

QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
 Db 2075 -----TTTCTCTTTGGTGGCACTATAACAGCCTTAATAGAGCTATTTCCTCA 2122
 QY 437 GluAsnGlyHisTrpLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
 Db 2123 ACTAAGGG-----GTTAAGCAAGCTCTGGTGGAGCGATTACAATCCAGGTTCTGTAT 2176
 QY 457 ThrAlaLeuIleArgThrSerAlaAlaGlyThrPhePheThrProGluAsnLysLys 476
 Db 2177 AACAAATACATAAACTAAGTGCAGATGTACAGGTTTCTACCCATTAGACAGATCAC 2236
 QY 477 LeuGlyThrPheIleLeuArgGlnAlaGlyThrValAlaAlaArgAspAsnAlaAsp 496
 Db 2237 CTCTGGGTTGTATCGCAAAAGCATCGCAGGATATGCAATGGTTTGTGAACAACGCGT 2296
 QY 497 ValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyThrGluLeu 516
 Db 2297 TTACCGGTTCTATCAAACTTATACAGCGGTGGCATTTGGTTTCATTACGCGGTTTGTCTAT 2356
 QY 517 AspSerIleGlyLeuAlaGlyProAsn----- 525
 Db 2357 GGTAGCAAT-----GGCCCTAACGCCAATTTATCAAGGTCAAAATAATAATTAAT 2407
 QY 526 -----GlySerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGln 542
 Db 2408 AAGATAAGTCTCATGATGATTGGTGAATGCAATCGTACAGCTAGCGCAGAGTTAAT 2467
 QY 543 LeuPro-----PheThrArgThr----- 548
 Db 2468 GTGCCAACTCCATTTGTGACTGATAGAGTCAAAATACAGTCGCAACCTCCCTATTGT 2527
 QY 549 -----LeuSerGlyAla 552
 Db 2528 GATCGCGCAAGTGTTCGAATACATAATGGAATACAGTAAATGATAGAGCAAT 2587
 QY 553 ValPheHisAspMetGlyAspAlaAlaAlaAsnPhelYsArgMetLysLeuLysHisGly 572
 Db 2588 GTCTTGAAGACATTACCGGATTTATGGC-----AAATCAAGCGCTACTCGCGCTCT 2638
 QY 573 SerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla--- 591
 Db 2639 ACAGGTGTCGGATTCCCAATGGCAATCTCCCTAGTGGACCACTGATTTTCTTATGCTAAA 2698
 QY 592 -----TyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThr 607
 Db 2699 CCAATTAATAATATGAAATGATGATGCTCAA---CAGTTCCAAATTTAGTTGGGGT 2755
 QY 608 ArgPhe 609
 Db 2756 TCCTTC 2761
 RESULT 34
 ID AAQ66199 standard; DNA; 2984 BP.
 XX AC AAQ66199;
 XX AC
 DT 08-DEC-1994 (first entry)
 DE H. Influenzae b Eagen D15 sequence.
 KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; Hib; ss.
 XX Haemophilus influenzae type b Eagen strain.
 OS
 FH Key Location/Qualifiers
 FT CDS 374..2767
 FT /*tag= a
 PN W09412641-A.

XX 09-JUN-1994.
 XX 23-NOV-1993; 93WO-CA00501.
 PR 23-NOV-1992; 92GB-0024584.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
 PI Yang Y;
 XX WPI: 1994-200269/24.
 DR P-PSDB; AAR53755.
 XX Nucleic acid encoding D15 outer membrane protein - esp. of
 PT haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 PS Disclosure; Fig. 1b; 16lpp; English.
 CC Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AAQ66198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 2984 BP; 960 A; 474 C; 619 G; 931 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,33e-09 Length: 2984
 Score: 209.50 Matches: 130
 Percent Similarity: 35.85% Conservative: 103
 Best Local Similarity: 20.00% Mismatches: 280
 Query Match: 6.60% Indels: 137
 DB: 15 Gaps: 24
 US-09-857-669-2 (1-609) x AAQ66199 (1-2984)
 QY 61 AspSerGluIleLysAspMetValGluHisLeuProLeuIleThrGlnGlnGlu 80
 Db 923 GAATCTGTTAGTAGCAGTACATTACAAGACAATGAATACACCTGATTCTTGCTGG 982
 QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
 Db 983 AAATTATGGGGAATAAATTTGAAGGTGCGCAATTCGAGAAGATTTGCAGTCAATTCGT 1042
 QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
 Db 1043 GATTATTATTAAATATATGCTATGCCAAGCACAAATTAATAAGCGGATGTTACAGCTA 1102
 QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
 Db 1103 AATGATGAAAAACAAAAGTTAATGTAACCATTCATGATCTAATGAGGTTTACAGTATGAC 1162
 QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
 Db 1163 CTTTCGTAGTCAGCGCATTTATAGGTAATCTGGGAGGTATGTCGCCGAGCTTGAACCTTTA 1222
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
 Db 1223 CTTTCAGCATTAATTAATGATCTTCCGCCGTAGT----- 1261
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
 Db 1262 -----GATATTGCAGATGTAGAAAAATGCAATTAAGCAAACTTGA--- 1303
 QY 186 ValThrArgLysGlyThrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 DB: ::::: ||| ::::: |||

Qy	526	-----GlySerValLeuProGlu 531
Db	2378	TATGGTAATGGTACTGCTACTGCTACTTTTAAAGAGATAAGTCTCTGATGCTGATGGTGGT 24337
Qy	532	ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db	2438	ANTGCAATCCGCTACAGCTAGCGCAGAGTTAATTGTGCCCACTCCATTGTGTGAGCGATTAAG 24977
Qy	545	PhetThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db	2498	AGCCAAAATACGTCCTGCGAACCTCCTTATTGTTGATCGCGCAAGTGTTTGGAATACTAAA 2557
Qy	558	-----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db	2558	TGGAATACAGATAAAATGGATTAGAGACGATGATTATTAAGAAAGATTGCCCTGATTATGGC 26177
Qy	568	-----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAla 584
Db	2618	AAATCAAGCGGTATTGCGCCCTCTACAGGTGCGGATTCCCAATGGCAATCTCCTATTGGG 26777
Qy	585	ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysile 599
Db	2678	CCATTGGTATTCTCTTATGCCAAACCAATTAAAAAATATGAAAATGATGATGTCGAA--- 2734
Qy	600	ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db	2735	CAGTTCCAATTATGATGAGGTTCTTTC 2764
RESULT 35		
AAQ66200	AAQ66200 standard; DNA; 2953 BP.	
AC	AAQ66200;	
XX	08-DEC-1994 (first entry)	
XX	H. influenzae b Minn A D15 sequence.	
DE	Vaccine; passive immunization; vector; antiserum; diagnosis; D15;	
KW	OMP; outer membrane protein; Hib; ss.	
XX	Haemophilus influenzae type b Minn A strain.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	CDS	334..2727
FT		/*tag= a
XX	W09412641-A.	
XX	09-JUN-1994.	
XX	23-NOV-1993; 93WO-CA00501.	
XX	23-NOV-1992; 92GB-0024584.	
XX	(CONN-) CONNAUGHT LAB LTD.	
PA	Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;	
XX	Yang Y;	
PI	WPI: 1994-200269/24.	
DR	P-PSDB; AAR53756.	
XX		
PT	Nucleic acid encoding D15 outer membrane protein - esp. of	
PT	Haemophilus influenzae, and related proteins, vectors, antisera	
PT	etc. useful in vaccines, for diagnosis and for passive	
XX	immunisation.	
XX	Disclosure; Fig. 1C; 16lpp; English.	
XX		
CC	Outer membrane protein (OMP) D15 genes were isolated by screening	
CC	chromosomal libraries of H. influenzae type b (Hib) strains Ca,	

CC Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AAQ66198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 CC The three 'n's are not identified in the specification.
 XX

50 Sequence 2953 BP; 948 A; 467 C; 619 G; 916 T; 3 other;

Alignment Scores:

Pred. No.: 2.85e-09 Length: 2953
 Score: 208.50 Matches: 130
 Percent Similarity: 35.85% Conservative: 103
 Best Local Similarity: 20.00% Mismatches: 280
 Query Match: 6.57% Indels: 137
 DB: 15 Gaps: 24

US-09-857-669-2 (1-609) x AAQ66200 (1-2953)

Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
 Db 893 GAATCTGTAGTACGATACATTTACAAGAACAAATGGAAATGACACCTGATCTTGGTGG 942
 Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
 Db 943 AAATATATGGGAAATAAATTTGAAGGTGCGCAATTCGAGAAAGATTTTCAGTCAATTCGT 1002
 Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
 Db 1003 GATTATTATTAATATATGCTATGCCAAAGCACAAATTAATAAACCGGATGTTTCAGCTA 1062
 Qy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
 Db 1063 AATGATGAAAAACAAAGTTAAATGCTAACCATTTGATGTAATGAAGTTTACAGTATGAC 1122
 Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
 Db 1123 CTTCGTAGTCGACGATTAAGTAATCTGGAGGTATGTCGCGAGCTTGAACCTTTA 1182
 Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 1183 CTTTCAGCATTAATTAATGATACTTTCGCGCGTAGT----- 1221
 Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
 Db 1222 -----GATATTCGACATGTAGAAATGCAATTAAGCAAACTTGA--- 1263
 Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 Db 1264 -----GAACGCGGTTAGCGGTAGCGCAACGCGTAAATTCAGTA-----CCT 1302
 Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
 Db 1303 GATTTTGATGATGTCAAAAACATTAAGGATAACCCCTGTGTGTGATGCTGGACGAGCT 1362
 Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
 Db 1363 TTAATGTTCCCAACTTCGCTTGAAGGAATACCGCTTTCGTGATAGCACTTTAGCT 1422
 Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
 Db 1423 CAGGAAATGCGCAACAAAGAGGACTTGGTATAATTCACAAATAGTTAGTTAGGAAAA 1482
 Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
 Db 1483 ATTGCTTAGATCGTACAGGTTTCTTCGAA-----ACAGTCGAAACCCGAATGATCCT 1536
 Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGlnValLysArgHis 299
 Db 1537 ATCAATGCTAGTAATGATGATGATGCTCGTATATAAAGTCAAGAACGTAACACGGT 1596
 Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319

Db 1597 AGTATCAACTTTGGTATTGGTTAGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGT 1656
 Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLys 339
 Db 1657 AAACAAGATAATTTCTTGGGAACAGGGGCGGAGTAAGTAGCTAGTGACGAAAAATGAT 1716
 Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
 Db 1717 TATGGTAGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTACTAAAGATGGTGTAGT 1776
 Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
 Db 1777 CTGGTGGAAATGTTTCTTTGAAAACTACGATAACTCTAAAGTGATACATCTCTTAAC 1836
 Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
 Db 1837 TATAAGCGTACGACTTAGCGGAAGTAATGTTACTTTTAGGTTTCCCTGTAATTAACAAATAC 1896
 Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
 Db 1897 TCCTATTATGTAGGTTAGGTACATCTATATAAATTAGTAATTTGCTCTAGATAT 1956
 Qy 403 -----GlyArgLysIleProGlySerAsp 410
 Db 1957 AACCGTAATTTATATCAATCAATGAATTAAGGTAATGCGATTTAAACAAATGAC 2016
 Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu 430
 Db 2017 TTTGAT-----TTTCTTTTGGTGGAACTATAACAGCGCTT 2052
 Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
 Db 2053 AATAGAGCGCTATTTCCCAACTAAAGG-----GTTAAAGCAAGCTTGGTGGAGAGT 2106
 Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
 Db 2107 ACTATTCAGGTTCTGTATAACAATACTACAACCTAAGTGCAGATGTACAGGGTTCTAC 2166
 Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyThrVal 490
 Db 2167 CCATTGACAGAGATCACCTCTGGGTGTATCTGCAAAAGCATCTGCAAGGATATGCAAT 2226
 Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
 Db 2227 GGTTTTGGAAACAAGCGTTTACCGTTCTATCAAACTTATACAGCGGTGGCATCGTTCA 2286
 Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
 Db 2287 TTAGTGGTGTGTTGTTTATGTTAGTATT-----GGACCTAACGCAATTTATGCCGAA 2337
 Qy 526 -----GlySerValLeuProGlu 531
 Db 2338 TATGGTAATGTAGTGGTACTGCTACTTTTAAAGAGATAAGTCTTGATGTGATTTGGTGGT 2397
 Qy 532 ArgAlaLeuLeuValGlySerLeuGlyTyrGlnLeuPro----- 544
 Db 2398 AATGCAATCGCTACAGCTAGCGAGAGTTAATTTGCGCAACTCCATTTGTGAGCGATAAG 2457
 Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
 Db 2458 ACCCAAAATACGGTCCGCAACCTCTCTATTGTTGTCGCGCAAGTGTGTTGGAATACTAAA 2517
 Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
 Db 2518 TCGAAATCAGATAAAATGGATTTAGAGAGCGGATGTTTAAAGAGATTCCTGATTTATGCG 2577
 Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
 Db 2578 AAATCAAGCCGTATTCGGCGCTCTACAGGTGTCGATTTCCCAATGCGCAATCTCCTATTGGG 2637
 Qy 585 ProPheSerPheIleAla-----TyrGlyHisSerAspLysIle 599

Db 2638 CCATTGGTATCTCTTATGTCACCAACCAATTAAAAATATGAAATATGATGATGTCGAA--- 2694
 QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
 Db 2695 CAGTTCCAATTTAGTATGAGGCTTCTTC 2724
 RESULT 36
 AAQ66198
 ID AAQ66198 standard; DNA; 2949 BP.
 AC AAQ66198;
 XX
 XX 08-DEC-1994 (first entry)
 DT
 XX
 DE H. influenzae b Ca D15 sequence.
 XX
 XX Vaccine; passive immunization: vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; Hib; ss.
 KW
 XX Haemophilus influenzae type b Ca strain.
 OS
 XX Key Location/Qualifiers
 FH 75..2468
 FT CDS /*tag= a
 FT
 XX W09412641-A.
 PN
 XX
 PD 09-JUN-1994.
 XX
 PF 23-NOV-1993; 93WO-CA00501.
 XX
 PR 23-NOV-1992; 92GB-0024584.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
 PI Yang Y;
 PI
 XX WPI; 1994-200269/24.
 DR P-PSDB; AAR53754.
 XX
 XX Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 PS Disclosure; Fig. 1A; 161pp; English.
 XX
 CC Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AA056198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 2949 BP; 983 A; 477 C; 612 G; 877 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.52e-09 Length: 2949
 Score: 207.50 Matches: 130
 Percent Similarity: 35.85% Conservative: 103
 Best Local Similarity: 20.00% Mismatches: 280
 Query Match: 6.54% Indels: 137
 DB: 15 Gaps: 24
 US-09-857-669-2 (1-609) x AAQ66198 (1-2949)
 QY 61 AspSerGluIleLysAspMetValGluHisLeuProLeuIleThrGlnGlnGlu 80
 Db 624 GAATCTGTTAGTACAGTACATTACAAGACAAATGGAATTACAACTGATCTTGGTGG 683

QY 81 GluValLeuAspLysGluInThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
 Db 684 AAATTATGGGAAATAAATTTGAAGTGGCAATTCGAGAAAGATTTGCAGTCATTCGT 743
 QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
 Db 744 GATTATTATTTAAATAATGGCTATGCCAAAGCACAATAATTACTAAACGGATGTCAGCTA 803
 QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProAlaGThrLys 133
 Db 804 AATGATGAAAAACAAACAAAGTTAATGTAACCATGATGTAATCAAGCTTTACAGTATGAC 863
 QY 134 -----TleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
 Db 864 CTTTCGTAGTCACGCATTTATAGTAAATCTGGGAGGTATGTCGCCGAGCTTGAACCTTTA 923
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 924 CTTTCAGCATTTACATTTAAATGATACTTCCGCCGTAGT----- 962
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
 Db 963 -----GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACATTGGA--- 1004
 QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 Db 1005 -----GAACGCGTTACGCTAGCGCAACGGTAAATTCAGTA-----CCT 1043
 QY 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
 Db 1044 GATTTGATGATGATCAAAATAAACATTAGCATAAACCTTCTTCTGATGCTGGACGCGT 1103
 QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
 Db 1104 TTAACCTGTTGCCCACTTCGCTTTGAAGAAATACCGCTTCTGCTAGTACACTTTACGT 1163
 QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
 Db 1164 CAGGAAATGCCCAACACANGAAGCACTTGTGTATAATTCACAATTAGTTAGTAGGAAAA 1223
 QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
 Db 1224 ATTTCGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAAAACCAATTGATCCT 1277
 QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
 Db 1278 ATCAATGGTAGTAATGATGAAGTGGATGCTGTATATAAAGTCAAGAACGTAACACGGGT 1337
 QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyr 319
 Db 1338 AGTATCAACTTTGGTATTTGGTTACGGTACAGACAGTGGTATTAGTTATCAAGCAAGTGTT 1397
 QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
 Db 1398 AAACAAGATAAATTTCTTGGACACAGGGCGGCAGTAAGTATAGCTGCTAGCAAAAATGAT 1457
 QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
 Db 1458 TATGGTACGAGTGTCAATTTGGTTATACCGAGCCTATTTTACTAAAGATGGTGTAAGT 1517
 QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
 Db 1518 CTTGGTGGAAATGTTTCTTTGAAACTACGATAACTCTAAAAGTGTATACATCCTCTAAC 1577
 QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
 Db 1578 TATAAGCGTACGACTTACGGAAGTAATGTTACTTTAGTTTCCCTGTAAATGAAAATAAC 1637
 QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
 Db 1638 TCCTATTATGTAGGATTAGGTATACCTATATAATAAATTAACCTTCTCTAGATAT 1697
 QY 403 -----GlyArgLysIleProGlySerAsp 410

```
Db 1698 AACGTAATTTATATATCAATCAATGAAATTTAAAGGTAATGCGATTAATAAACAATGAC 1757
Qy 411 ileAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTTTCTTTTGGTGGAACTATAACAGCGCTT 1793
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTrpLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 AATAGAGCGTATTTCCCACTAAAGG-----GTTAAAGCAAGCTGTGGTGACGAGTT 1847
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCAGCGTTCTGATAACAATACTACAACAACTAAGTCAGATGTACAGGGTTTCTAC 1907
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAAAT 1967
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTTTTGGAACACAGCGTTTACCGTTCTATCAAACTTATACAGGGGTGCATCGGTCA 2027
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGGTTTCTCTTATGCTAGTATT-----GGACCTAACGCAATTTATGCCGAA 2078
Qy 526 -----GlySerValLeuProGlu 531
Db 2079 TATGGTAATCGTACTGCTACTGGTACTTTTAAAGAAAGATAAGTCTGATGTGTTGGTGT 2138
Qy 532 ArgAlaLeuLeuValGlySerLeuGlyTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCCGAGAGTTAATTTGTGCCAACTCCATTGTGAGCGATAAG 2198
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisaspMet----- 557
Db 2199 AGCCAAATACGGTCCGAAACCTCTTATTTGTTGATCGCGCAAGTGTGGAAATACTAAA 2258
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGGATTAGACGCGATGTATTAAGAAAGATTGCGCTGATTATGGC 2318
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AATCAAGCGGTATTCGCGCTCTACAGGTGTCGAGTTCCCAATGCCAATCTCTATTGGG 2378
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGTATTCTCTTATGCGCAACCACTAAATAATATGAAATGATGATGTCGAA--- 2435
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCAATTTAGATTGGAGGTTCTTTC 2465

RESULT 37
AAF61065
ID AAF61065 standard; DNA; 2019 BP.
XX
AC AAF61065;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF08500.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE1935088-Al.
XX
PD 01-FEB-2001.
```

```
XX
XX 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GIBF) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT. KREBSFORSCHUNGSZENTRUM.
PA (WEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX WPI: 2001-192469/20.
DR
XX
XX New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria.
PS
PS Claim 1a: Page 117-118; 158pp; German.
XX
XX This invention describes novel DNA sequences (I) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (1); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (1) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (1) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
CC carrying one or more (1). (1), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (1),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 2019 BP; 404 A; 671 C; 603 G; 338 T; 3 other;

Alignment Scores:
Pred. No.: 7.59e-09 Length: 2019
Score: 201.50 Matches: 110
Percent Similarity: 34.36% Conservative: 57
Best Local Similarity: 22.63% Mismatches: 224
Query Match: 6.35% Indels: 95
DB: 22 Gaps: 19

US-09-857-669-2 (1-609) x AAF61065 (1-2019)
Qy 180 ThrSerValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLys 195
Db 70 ACCACACGTCGCGAGCTGATCACCCTCCGCGCTGGGTAAAGAGCTACCTCGCTAAC 129
Qy 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnVal 215
Db 130 GTC---AACGGCGTGGCGCAACCCACACGACGAGCCACACGCTGACATCATGTTCTGTG 186
Qy 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlnArgTyr 235
Db 187 CTCACCGCGGCAAGCGTGGCTACGTCAACGCGCTACCACTACCGCGCAACCAAGACC 246
Qy 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 247 GAAGACCGAGTGTGCTGCGTCCGGAATCGCCASATGAAGCGCTGGCGCTCACCTAC 306
Qy 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerVal 275
```

Db	1327	GGCAACGTGTATGCCGACACCTGCTACTGTCACACCCAAAGCGCTGGGGCAGCGTGCAC	1386
Qy	574	-----GlyLeuGlyValArgTrrpPheSerProLeuAlaPro	585
Db	1387	CTGGCGCAGATGGCTGTGTGCTGGGGTGGCGTACCTGGTACAGCCGATGGGGCCG	1446
Qy	586	PheSerPheAspIleAla	591
Db	1447	CTGAGTTTCAGCCTGGCG	1464
RESULT	38		
AAV25138			
ID	AAV25138	standard; DNA; 2778 BP.	
XX	AC		
XX	AC		
XX	XX		
XX	XX		
XX	13-JUL-1998	(first entry)	
XX	H. pylori	cell envelope OMP ORF 05ep10815_4719175_c1_l15.	
DE	XX		
XX	Cytoplasmic; vaccine; prevention; treatment; infection; envelope;		
KW	identification; binding compound; bacteria; life cycle; activator;		
KW	inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;		
KW	cell envelope outer membrane protein; OMP; ds.		
XX	XX		
XX	OS	Helicobacter pylori.	
XX	XX		
PH	Key	Location/Qualifiers	
FT	CDS	1...2778	
FT		/*tag= a	
XX	XX		
XX	W09737044-A1.		
XX	XX		
XX	09-OCT-1997.		
XX	XX		
XX	27-MAR-1997;	97WO-US05223.	
XX	06-DEC-1996;	96US-0761318.	
PR	29-MAR-1996;	96US-0625811.	
PR	02-APR-1996;	96US-0758731.	
PR	25-OCT-1996;	96US-0736905.	
PR	28-OCT-1996;	96US-0738859.	
XX	XX		
PA	(ASTR)	ASTRA AB.	
XX	XX		
PI	Alm RA,	Smith D;	
XX	XX		
DR	WPI;	1997-503122/46.	
DR	P-PSDB;	AAW55729.	
XX	XX		
PT	Helicobacter pylori	nucleic acid sequences and encoded	
PT	polypeptide(s)-	useful in vaccines to treat or prevent H. pylori	
PT	infection and for	diagnosis of H. pylori infection	
XX	Claims	5,6,21; Page 929,930; 1145pp; English.	
PS	XX		
CC	XX	This sequence encodes a H. pylori cell envelope outer membrane	
CC	XX	protein having a terminal phe residue.	
CC	XX	The protein may be used in a vaccine to prevent or treat H. pylori	
CC	XX	infection or to identify H. pylori polypeptide binding compounds,	
CC	XX	useful as potential H. pylori life cycle activators or inhibitors. The	
CC	XX	DNA and probes derived from it may be used for the identification of	
CC	XX	H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic	
CC	XX	acid sequences complementary to the DNA act as antisense sequences and	
CC	XX	can be used to prevent the translation of H. pylori mRNA. Antibodies	
CC	XX	against the protein can be used in immunoassays to evaluate the abundance	
CC	XX	and distribution of H. pylori-specific antigens. The genomic sequence of	
CC	XX	H. pylori (ATCC 55679) was determined from overlapping contigs generated	
CC	XX	by mechanically shearing the bacterial DNA. The sequences were analysed	
CC	XX	for ORF of at least 180 nucleotides, and the predicted coding regions	
CC	XX	defined by computer evaluation. To identify likely H. pylori antigens for	
CC	XX	vaccine development, the amino acid sequences predicted from various ORF	
CC	XX	were analysed for significant homology to other known or exported	

CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from *H. pylori* by PCR
CC amplification for recombinant polypeptide production, e.g. in *E. coli*
CC hosts.
XX
SQ Sequence 2778 BP; 813 A; 514 C; 681 G; 770 T; 0 other;

Alignment Scores:

Pred. No.:	4.25e-06	Length:	2778
Score:	174.00	Matches:	126
Percent Similarity:	33.67%	Conservative:	72
Best Local Similarity:	21.43%	Mismatches:	197
Query Match:	5.48%	Indels:	194
DB:	18	Gaps:	23

US-09-857-669-2 (1-609) x AAV25138 (1-2778)

Qy	186	ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro	205
Db	1102	ATCCCGGATAGGGCTATGCGTTT-----GCGGTGGTGAAGCCA	1140
Qy	206	AspThrAlaThrValAspLeuAsnValVal-----AspSerGly	219
Db	1141	GACTTGGATAAGACGAAAMACGGCTTGTGAAGTCATTTATCGTATTGAAGTGGC	1200
Qy	220	ArgProIleAlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIle	239
Db	1201	GATATGGTGATATCAATGATGTCATCTTCAGGAACACGCGCAGCGGATAGATC	1260
Qy	240	ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp	259
Db	1261	ATTAGGAGGAATTTGTACTAGGCGCTAAAGATAAATACAACTTGACCAAACTGAGAAAT	1320
Qy	260	PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe	279
Db	1321	TCCGAAATATCTTTAGGCGGTTTAGGTTTCTCT-----AAAGTCAAGATTGAAGAA	1374
Qy	280	AspArgLeuGlnGlnGlyAspArgValProValLysValSerValThrGlnValLysArgHis	299
Db	1375	AAAAGGTCAATACCTCATTGATGATTTGTTAGTACGCTAGAGAGGGCGGACCCGG	1434
Qy	300	LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr	319
Db	1435	CAGTTGAATTCGGGTTGGGCTATGGCTCTTATGAGGCGCTCATGCTTAATGGAGCGTG	1494
Qy	320	AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAsp-----	336
Db	1495	AGCGAAGGAATCTTTTGGCACAGGCGAAGCATGACCTTGATGCTACATTTGCCACA	1554
Qy	337	-----MetAspLysTyrGluThrThrLeuAlaAlaGly---	347
Db	1555	GGGGGGGTAGATCTTATCCGGGCATGCCAAAGGGCGGGCGCTATGTTTGGCGGGAAT	1614
Qy	348	-----IleSerGlnProArg-----	352
Db	1615	TTGAGCTTGACATAATCCAAGGATTTTTCACAGCTGGTATAGCTCTACGATCAATCTTAT	1674
Qy	353	---AsnTyrArg-----	355
Db	1675	CGGATTACAGGATTAAGCTACCAATACATCAACAAGCGGGGGCTTTGGGGTGAATGTC	1734
Qy	356	-----GlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln	371
Db	1735	GGGGCATGCTGGGTAATAGAACCCATGTGAGCTTAGGGTATAACTTCAATGTTACCAAA	1794
Qy	372	AsnLeuGluLysArgAlaPheSerGlyIleTyrTyrValArgAspArgAlaGlyIle	391
Db	1795	CTCCTT-----GGTTTCAGCAGCCCTTA---TACAAACCGCTACTATTCTCTGTT	1842
Qy	392	Asp-----AlaArgLeuGlyAlaGluPheLeu	400
Db	1843	AATGAAGTGGCTCTCCAAAGCAATGTTCCACACCGCATCGGTGATTTCAACCGCTTA	1902

Qy	401	AlaGluGlyArgLys-----IleProGlySerAspIleAspLeuGly-----	414
Db	1903	TCAGGCGGTAGAACTCCATTGGTTCTGAAAGCTGTTCTAGTCTCTGGAGCGATCACCAC	1962
Qy	415	-----AsnSerHisAlaThrMetLeuThrAlaSerTrp	425
Db	1963	TCACCAGAAATAAAAGGTATTTGGGATAGGATTTACCACAGCGCTATCACCAGTCTTTC	2022
Qy	426	LysArgGlnLeu-----LeuAsnAsnValLeuHisProGluAsnGlyHisTyr	441
Db	2023	ACCCTTGATGTGAGCTATGACACACCGATGATTTATTTCCCTAGAAATGGGGTTATC	2082
Qy	442	LeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeu-IleAr	461
Db	2083	TTTAGTTCCCTAT-----GCGACAATGCTCTGGTTTGCACAGCTCTGCGCAGCTCAATCT	2136
Qy	461	gThrSerAlaArgAlaGlyTyrPhePheThrProGlu-----AsnLysLysLeuG	478
Db	2137	TGGAACGGGTTAGCGGGAATGTCCTTAACCAAAAGTTTATGTAATTTGCGCGCTTAC	2196
Qy	478	yThr-----PheIleIleArgGlyGlnAlaG	487
Db	2197	CACCATTTGC AAAAATATTTATGTAGATGTTGATCGCTCGTTTAAAACG-CAAGGGGG	2255
Qy	487	yTyrThrValAlaArgAspAsnAlaAsp-----ValProSerGlyLeuMetPheArgSe	505
Db	2256	CTAT---ATCTTTAGGTATAACACCGATGATTACTTGGCCTTAAACTCCACTTCTACAT	2312
Qy	505	rGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAs	525
Db	2313	GGGGGGCTAACCCAGCTGAGAGCTTTAGG-----AA	2345
Qy	525	nGlySerValLeuProGlu-----	531
Db	2346	CGGCTCAATCACACCTAAAGATGATGTTGGCTTGTGGAGCGCATGGGATTTTTTAC	2405
Qy	531	-----	531
Db	2406	CGCTTCTACTCAATTGAGCTATGGGTGTTAAAGCGGCTAAATGCGTTTAGCGTGGTT	2465
Qy	532	-----ArgAlaLeuLeuValGlySerLeuGluTyrGlnLe	543
Db	2466	TTTTGACTTTGGTTTCTTAACCTTTAAACCCCAACTAGGGGAGTTTCTTCTATAACGC	2525
Qy	543	uProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAs	563
Db	2526	TCCACACACGACG-----GCGAATTTTAAAGATATGCGCTGTAGGGGCTGG	2573
Qy	563	nPheLysArgMetLysLeuLysHisGlySerGlyLeuValArgTrpPheSerProLe	583
Db	2574	GTTTGAAGGGCGACTTTGGAGGGCTTCTACAGGCTTACAGATTCATGATGGATTCGCCAT	2633
Qy	583	uAlaPro-----PheSerPheAspIleAlaTyr-----	592
Db	2634	GGGCGCTTGGTGTGATTTTCCCTATAGCGTTTTTCAACCAATGGGCGCATGCAATGG	2693
Qy	593	-----GlyHisSerAspLysLysIleArgTrpH	602
Db	2694	CAAAAAATGAAAGGGCTGTGCTTTTAAACCTTAACCTTAACGATGATGATGATGATGATG	2753
Qy	602	sIleSerLeuGlyThrArgPhe	609
Db	2754	ATTTTCTATGGGAACAAGGTTT	2775

RESULT 39

AAV75802

ID AAV75802 standard; DNA; 2778 BP.

XX

AC AAV75802;

XX

DT 03-AUG-1999 (first entry)

XX

DE *H. pylori* outer membrane polypeptide encoding DNA.

XX	Outer membrane polypeptide; OMP; vaccine; <i>H. pylori</i> infection; humoral;
KW	cellular immune response; ds.
XX	
OS	<i>Helicobacter pylori</i> .
XX	
PN	WO9921959-A2.
XX	
PD	06-MAY-1999.
XX	
PD	28-OCT-1998; 98WO-US22883.
XX	
PF	17-DEC-1997; 97US-0993001.
PR	28-OCT-1997; 97US-0959131.
XX	
PR	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Alm RA, Ellis RW, Gulid BC, Noonan BM, Smith D;
XX	
DR	WPI; 1999-326698/27.
XX	P-PSDB; AAY17183.
XX	
PT	Cellular vaccine against <i>Helicobacter pylori</i>
XX	
PS	Claim 8; Page 147-149; 352pp; English.
XX	
CC	The invention relates to a vaccine for preventing or treating infections
CC	by <i>Helicobacter pylori</i> . The vaccine contains at least one isolated
CC	<i>H. pylori</i> polypeptide, or its fragments, in a carrier, where the
CC	carrier is a <i>Salmonella</i> , <i>Vibrio cholerae</i> or <i>Shigella</i> vector containing a
CC	nucleic acid encoding the <i>H. pylori</i> polypeptide. The vaccines induce
CC	humoral and cellular immune responses. The vaccines are used to treat or
CC	prevent infections by <i>H. pylori</i> . Sequences AAX75779 to AAX75837 represent
CC	nucleic acid sequences encoding <i>H. pylori</i> outer membrane polypeptides
CC	(OMPs) AAY17160 to AAY17218.
XX	
SQ	Sequence 2778 BP; 813 A; 514 C; 681 G; 770 T; 0 other;

```

Alignment Scores:
Pred. No.: 4,25e-06 Length: 2778
Score: 174.00 Matches: 126
Percent Similarity: 33.67% Conservative: 72
Best Local Similarity: 21.43% Mismatches: 197
Query Match: 5.48% Indels: 194
DB: 20 Gaps: 23

US-09-857-669-2 (1-609) x AAX75802 (1-2778)

Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
   :::: |||||:::
Db 1102 ATCCCGCATAGGCGTCATCGGTT-----GGGTGTGGGAAGCCA 1140

Qy 206 AspThrAlaLehrValAspLeuAsnValVal-----AspSerGly 219
   ||| :::: ||| ::::
Db 1141 GACTTCGATAAACACGAAAAACGGCGTGTGAAAGTCATTTATCGTATTGAAGTGGC 1200

Qy 220 ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle 239
   :::: ||| :|||::: ||||| ::::
Db 1201 GATATGGTGTCATTCATCATGATGTCATCATTTTCAGGGAACACGCGCAGCGGATAGGATC 1260

Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
   :::: ||| :|||::: ||| ::::
Db 1261 ATTAGGAGGGAATTGTTTACTAGGCGCTAAGAGATAAATACAACCTGACCAAACTCAGAAAT 1320

Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
   :::: :|||::: ||| :|||::: |||:::
Db 1321 TCCGAAATTCCTTTGAGCGGCTTTAGGGTTTTCTCT-----AAAGTCAAGATTGAAGAA 1374

Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
   |||::: :|||::: ||| :|||::: ||| ::::
Db 1375 AAAGGTCGAATAGCTCATGTATGGATTGTTAGTAGCGGTAGAAAGGGCGGCACCGGG 1434

Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319

```


Db 745 GCTGATGCTATA-----GTCAACTCTCACTATGACCTTGACGACAAA 786
Qy 176 GluAsn-----SerLysThrSerValLeuGlyAlaVal 186
Db 787 GGGAAATATCTCTTTACATGGATATTGATCGAGGGTCGGCATATACCTTAGGACACGTC 846
Qy 187 ThrArgLysGlyTyrProLeu----- 193
Db 847 CATATCCAAGGGTTTGAGGTTTCCCAAAAGCCCTTATAGAAAAGCAATCCCAAGTCGC 906
Qy 193 ----- 193
Db 907 CCCAATGATCTTTATGCCCGGATAAATATGGGATGGGCTCATAGATCAAAACAACT 966
Qy 194 ---AlaLysLeuGly-----AsnThrArgAlaAlaVal-----AsnProAsp 206
Db 967 TATGCAAGATGCTACATCAATACCAATGTACAGCTTCTCTTCATCCCTCAGCGCAACC 1026
Qy 207 ThrAlaThrValAsnValValValAspSerGlyArgProIleAlaPheGlyAsp 226
Db 1027 CGCCCTATTTATGATTAACCTTATGAGTAAGTGAAGGCTCTCCTTATAAAGTTGGGTTA 1086
Qy 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
Db 1087 ATTAATAATTTACTGGGAATACCACAAATCTCACGTTATTTTACACGAAACCACTCTC 1146
Qy 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 1147 TTCCAGGAGATACATCAATCGCTTAAGCTAGAAGTACTGAGCAACGTTTAAAGAAAT 1206
Qy 267 AsnGlyHisTyrSerGlyAlaSerVal-----GlnAla 277
Db 1207 ACAGCTACTTCCAAAGGGTTAGTGCTCTATACAGTTCGTTCTCACTTGATCTATGGC 1266
Qy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGlu----- 295
Db 1267 AATCGGATCAATACCGAGATATTTTGTAGAAGTCAAAAGAAACACACAGAAACTTA 1326
Qy 296 -----ValLysArgHisLysLeuGluThrGlyIleArgLeu---Asp 308
Db 1327 GCCTTATCTTAGGATTTAGTCTCTTGACAAATCTTTTGGAGGAATTTGAACCTATCTGAA 1386
Qy 309 SerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGly 328
Db 1387 AGTAATTTTGATCTATTGGAGCT-----AGAAATATATTTTCTAAAGGT 1431
Qy 329 Tyr-----IleGlySerVal 333
Db 1432 TTTGCTTCTTAAGAGCGGTGGAGACATCTATTTTAAAGGCAACTTCGGGGACAAA 1491
Qy 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
Db 1492 GTCACAGACTATCTTGAAGTGGACC-----AAACCT---CAT 1527
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnVal-----SerTyrAsnArgSerThr 369
Db 1528 TTTCTAAACACTCTCTGGATTTAGGAATTCGAATAGATAATCAATTAACAGACCATTA 1587
Qy 370 ThrGlnAsnLeuGluLysArgAlaPheSerGlyIle-----TrpTyrValArgAsp 387
Db 1588 TCTAAGATTTATGTGTCCAAACCTATGGCGGAACGTCAGCACAACCTATATCTTGAAC 1647
Qy 388 -----ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArg 404
Db 1648 GAACACCTGAATACGGTCTATTTTATCGAGGAAGTCAACAGGTTTACATGAAGAAACGCT 1707
Qy 405 Lys-----IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
Db 1708 AAGTCTCTCTAGGCGCAAAATATAGAC-----ACCAATAAAGGATTTGTCTCTGCT 1758
Qy 424 SerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrIleuAsp 443
Db 1759 GCAGGTGTCAACTTGAATTACGATTTCTGTAGATAGTCTCTAG-ACCTCCAACCTACAGGGAT 1817

Qy 444 GlyLysIleGlyThrThrLeuGlyThrPhe-LeuSerSerThrAlaLeuIleArgThrse 463
Db 1818 TCGGGGGGGGTGACTTTTGGAGTTTCTGGTTGGGAGGAAC----- 1860
Qy 463 rAlaArgAlaGlyTyrPhePheThrProGluAsn-----LysLysLe 477
Db 1861 -----TATCATTTTACAAAACCTCTCTTTAAACAGCTCTATCTATAGAAAAC 1907
Qy 477 u-----GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaAArgAspAs 494
Db 1908 TACCGGTAAAGGTATTTGAAAATCAAGGGGAAGCTCAATTTTAAACCTTATAGCAA 1967
Qy 494 nAla-----AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerVa 511
Db 1968 TACTACAGCTGAAGGAGTCTCTGTCACTGAGCGCTTCTCTCTAGGTGGAGAGACTACAGT 2027
Qy 511 lArgGlyTyrGluLeuAspSerIleGly-----LeuAlaGlyProAsnGlyse 527
Db 2028 TCGGGGATATAAATCCTTTATATCGGTCCAAAATACTCTGCTACAGAACTCAGGAGG 2087
Qy 527 rValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrAr 547
Db 2088 ACTC-----TCTTCGCTCCTTATT---TCAGAAAGGTTTCAATACCCCTCTCATCAG 2135
Qy 547 q-----ThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLy 565
Db 2136 ACAACCTTAATATTAGTCCCTTTGTATTCTTAGACTCAGTTTTCGCGTTTACAGAGTA 2195
Qy 565 sArgMetLysLeuLys-----HisGlySerGlyLeuGlyValArgTrp 579
Db 2196 TAAGATTTTCGTTAAAGATCTACGTAGTAGTGGTGGATTTGGTCTCGCTTC 2247

Search completed: November 9, 2002, 01:37:16
Job time : 726 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:13:49 ; Search time 35 Seconds
(without alignments)
2318.561 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3173	100.0	609	21 AAY99623	Neisseria meningit
2	3173	100.0	609	21 AAY99624	Neisseria meningit
3	3153	99.4	615	21 AAY74862	Neisseria meningit
4	3136	98.8	615	21 AAY74861	Neisseria meningit
5	3019	95.1	587	21 AAY99625	Neisseria meningit
6	1443	45.5	522	21 AAY74860	Neisseria gonorrhoe
7	374	11.8	578	21 AAY95820	Haemophilus influe
8	355.5	11.2	576	21 AAY95821	Haemophilus influe
9	260.5	8.2	792	21 AAY84946	Amino acid sequenc
10	256.5	8.1	792	22 AAU03958	Neisseria meningit

11	256.5	8.1	792	22 AAB84745	Amino acid sequenc
12	256.5	8.1	792	22 AAB23786	Neisseria gonorrhoe
13	253	8.0	797	22 AAU03957	Neisseria meningit
14	253	8.0	797	22 AAU03959	Neisseria gonorrhoe
15	253	8.0	797	22 AAU04451	Neisseria meningit
16	253	8.0	797	22 AAB84744	Amino acid sequenc
17	253	8.0	797	22 AAB84746	Amino acid sequenc
18	253	8.0	797	22 AAB23784	Neisseria meningit
19	253	8.0	797	22 AAB23788	Neisseria meningit
20	251	7.9	797	21 AAY84947	Amino acid sequenc
21	247	7.8	896	22 AAG78603	Lawsonia intracell
22	222.5	7.0	474	22 ABG17710	Novel human diagno
23	219	6.9	889	21 AAY95988	Moraxella catarrha
24	219	6.9	919	21 AAY95987	Moraxella catarrha
25	213.5	6.7	793	15 AAR53758	H. influenzae PAK
26	210.5	6.6	792	15 AAR53757	H. influenzae b Pa
27	209.5	6.6	797	15 AAR53755	H. influenzae b Pa
28	208.5	6.6	797	15 AAR53756	H. influenzae b Pa
29	207.5	6.5	797	15 AAR53754	H. influenzae b Pa
30	174	5.5	925	18 AAW55729	H. pylori ORF 05ep
31	174	5.5	925	20 AAY17183	H. pylori outer me
32	173.5	5.5	790	23 ABB90599	Chlamydia pneumoni
33	173.5	5.5	795	20 AAY34896	Chlamydia pneumoni
34	172.5	5.4	847	19 AAW71477	Helicobacter poly
35	172.5	5.4	916	22 AAB46311	H. pylori HPS120 p
36	169.5	5.3	787	21 AAY69362	Amino acid sequenc
37	165.5	5.2	813	21 AAY44390	M. catarrhalis BAS
38	165.5	5.2	813	21 AAY44391	M. catarrhalis (AT
39	164	5.2	797	20 AAY36955	Chlamydia trachoma
40	160.5	5.1	371	22 ABG25558	Novel human diagno
41	150	4.7	818	21 AAG32160	Arabidopsis thalia
42	150	4.7	818	23 ABB92565	Herbicidally activ
43	149.5	4.7	891	20 AAY34495	Porphyromonas ging
44	149.5	4.7	899	20 AAY34369	Porphyromonas ging
45	146	4.6	3472	21 AAY90913	Cenarchaeum symbio

ALIGNMENTS

RESULT 1
AAY99623
ID AAY99623 standard; Protein: 609 AA.

AC AAY99623;

DT 03-JAN-2001 (first entry)

XX Neisseria meningitidis BASB040 putative protein sequence #1.

DE BASB040: bacterial disease; respiratory tract infection; bacteraemia;

KW meningitis; cancer; autoimmune disease.

XX Neisseria meningitidis.

PN WO2000034480-A1.

PD 15-JUN-2000.

XX 02-DEC-1999; . 99WO-EP09560

PR 07-DEC-1998; 98GB-0026886.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

DR WPI: 2000-423426/36.

XX N-PSDB; AAA48507.

Novel BASB040 polypeptides of Neisseria meningitidis useful for

diagnostic, prophylactic and therapeutic purposes against microbial

PT diseases comprise a specific amino acid sequence

12/2/77
12/7/95

```
XX PS Claim 4; Page 59-60; 98pp; English.
XX CC The present sequence is a putative version of the Neisseria
XX CC meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
XX CC similar to the D15 outer membrane protein of the bacterium. The protein,
XX CC its gene, antibodies, antagonists and agonists can be used to diagnose
XX CC and treat bacterial diseases such as those leading to upper respiratory
XX CC tract infections, bacteraemia and meningitis. In addition, they can be
XX CC used in vaccines for use against cancer and autoimmune diseases.
XX SQ Sequence 609 AA;
Query Match 100.0%; Score 3173; DB 21; Length 609;
Best Local Similarity 100.0%; Pred. No. 8.2e-252;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MMKPTALLPALFFPHAYAPAAADSENKAAGPALFNKNSPDTSVKLPKFPVRIDTQ 60
Db 1 MMKPTALLPALFFPHAYAPAAADSENKAAGPALFNKNSPDTSVKLPKFPVRIDTQ 60
Qy 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGFLAEAPONVKTMLRSKGYFSSKVSLETKDG 120
Db 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGFLAEAPONVKTMLRSKGYFSSKVSLETKDG 120
Qy 121 AYTTHITPGPRTKIANYGVAITLGDILSDGNLAEEYRNALENWQOPVGSDFDQDSWENSKT 180
Db 121 AYTTHITPGPRTKIANYGVAITLGDILSDGNLAEEYRNALENWQOPVGSDFDQDSWENSKT 180
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Qy 241 SGLARFOPGTPYDLDLDDFQQALEQNGHYSGASVQADFQRLQGDPRVPKVSVEVRHK 300
Db 241 SGLARFOPGTPYDLDLDDFQQALEQNGHYSGASVQADFQRLQGDPRVPKVSVEVRHK 300
Qy 301 LETGIRLDSYGLGCKIAYDYNNLFNKGYIGSVVWMDKYETTLLAAGISQPRNRYGNWT 360
Db 301 LETGIRLDSYGLGCKIAYDYNNLFNKGYIGSVVWMDKYETTLLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTTQNLKRAFSGGIWYVRORAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTTQNLKRAFSGGIWYVRORAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Qy 481 IIRQAGYTVARDNADVPGLMFRSGGASSYRGYELDSIGLAGPNSVLPERALLVGSLE 540
Db 481 IIRQAGYTVARDNADVPGLMFRSGGASSYRGYELDSIGLAGPNSVLPERALLVGSLE 540
Qy 541 YQLPFTRTLGSVPHDMDAAANPKRMKLKHGSLGVWRFSPAPFSDIAYGHSDKKIR 600
Db 541 YQLPFTRTLGSVPHDMDAAANPKRMKLKHGSLGVWRFSPAPFSDIAYGHSDKKIR 600
Qy 601 WHISLGTRF 609
Db 601 WHISLGTRF 609
RESULT 2
AAY99624
ID AAY99624 standard; Protein; 609 AA.
AC AAY99624;
XX 03-JAN-2001 (first entry)
XX Neisseria meningitidis BASB040 putative protein sequence #2.
DE
XX
XX BASB040; bacterial disease; respiratory tract infection; bacteraemia;
```

```
KW PS Claim 4; Page 60-61; 98pp; English;
XX CC The present sequence is a putative version of the Neisseria
XX CC meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
XX CC similar to the D15 outer membrane protein of the bacterium. The protein,
XX CC its gene, antibodies, antagonists and agonists can be used to diagnose
XX CC and treat bacterial diseases such as those leading to upper respiratory
XX CC tract infections, bacteraemia and meningitis. In addition, they can be
XX CC used in vaccines for use against cancer and autoimmune diseases.
XX SQ Sequence 609 AA;
Query Match 100.0%; Score 3173; DB 21; Length 609;
Best Local Similarity 100.0%; Pred. No. 8.2e-252;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MMKPTALLPALFFPHAYAPAAADSENKAAGPALFNKNSPDTSVKLPKFPVRIDTQ 60
Db 1 MMKPTALLPALFFPHAYAPAAADSENKAAGPALFNKNSPDTSVKLPKFPVRIDTQ 60
Qy 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGFLAEAPONVKTMLRSKGYFSSKVSLETKDG 120
Db 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGFLAEAPONVKTMLRSKGYFSSKVSLETKDG 120
Qy 121 AYTTHITPGPRTKIANYGVAITLGDILSDGNLAEEYRNALENWQOPVGSDFDQDSWENSKT 180
Db 121 AYTTHITPGPRTKIANYGVAITLGDILSDGNLAEEYRNALENWQOPVGSDFDQDSWENSKT 180
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Qy 241 SGLARFOPGTPYDLDLDDFQQALEQNGHYSGASVQADFQRLQGDPRVPKVSVEVRHK 300
Db 241 SGLARFOPGTPYDLDLDDFQQALEQNGHYSGASVQADFQRLQGDPRVPKVSVEVRHK 300
Qy 301 LETGIRLDSYGLGCKIAYDYNNLFNKGYIGSVVWMDKYETTLLAAGISQPRNRYGNWT 360
Db 301 LETGIRLDSYGLGCKIAYDYNNLFNKGYIGSVVWMDKYETTLLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTTQNLKRAFSGGIWYVRORAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTTQNLKRAFSGGIWYVRORAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Qy 481 IIRQAGYTVARDNADVPGLMFRSGGASSYRGYELDSIGLAGPNSVLPERALLVGSLE 540
Db 481 IIRQAGYTVARDNADVPGLMFRSGGASSYRGYELDSIGLAGPNSVLPERALLVGSLE 540
```

Db 481 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGLVPERALLVGSLE 540
QY 541 YQLPFTTLTSGAVFDMGDAANFKRMKLGSGLVGRWFSPLAPFSDIAYGHSKKIR 600
Db 541 YQLPFTTLTSGAVFDMGDAANFKRMKLGSGLVGRWFSPLAPFSDIAYGHSKKIR 600
QY 601 WHISLGRF 609
Db 601 WHISLGRF 609
RESULT 3
AA74862
ID AAY74862 standard; Protein; 615 AA.
AC AAY74862;
XX 21-MAR-2000 (first entry)
DT Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1198.
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
(CHIR) CHIRON CORP.
(GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX WPI: 2000-062150/05.
DR N-PSDB: AA253624.
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX Claim 2; Page 669; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX Sequence 615 AA;

Query Match 99.4%; Score 3153; DB 21; Length 615;
Best Local Similarity 99.3%; Pred. No. 3.7e-250;
Matches 605; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MMKPTALLPALPFFPHAYAPADLSENKAAGPALFKNSPDPTESVKLPKFPVPRDITQ 60
Db 7 MMKPTALLPALPFFPHAYAPADLSENKAAGPALFKNSPDPTESVKLPKFPVPRDITQ 66
QY 61 DSEIKDMVEHPLITIQOQEVLDKEOTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG 120
Db 67 DSEIKDMVEHPLITIQOQEVLDKEOTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG 126
QY 121 AYTTHITPGPRTKIANYGVAILGSDGNLAEYRNALENWQOPVGSDFDQDQSWENSKT 180
Db 127 AYTTHITPGPRTKIANYGVAILGSDGNLAEYRNALENWQOPVGSDFDQDQSWENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIARGDEITGTQRYPSQIV 240
Db 187 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIARGDEITGTQRYPSQIV 246
QY 241 SGLARFPGTGYDLDLDFQOALEQNGHYSGASVQADFRLQGDVPRVKVSVTEVKRHK 300
Db 247 SGLARFPGTGYDLDLDFQOALEQNGHYSGASVQADFRLQGDVPRVKVSVTEVKRHK 306
QY 301 LETGIRLDSEYGLGKIAYDYNNLFNKGYIGSVVWMDKYEYTTLAAGISQPRNRYGNYWT 360
Db 307 LETGIRLDSEYGLGKIAYDYNNLFNKGYIGSVVWMDKYEYTTLAAGISQPRNRYGNYWT 366
QY 361 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 426
QY 421 LTASWKKQLNNVLPENHYLDGKIGTTLGTLSSLTALINTSARAGYFFFTPEKKLGTF 480
Db 427 LTASWKKQLNNVLPENHYLDGKIGTTLGTLSSLTALINTSARAGYFFFTPEKKLGTF 486
QY 481 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGLVPERALLVGSLE 540
Db 487 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGLVPERALLVGSLE 546
QY 541 YQLPFTTLTSGAVFDMGDAANFKRMKLGSGLVGRWFSPLAPFSDIAYGHSKKIR 600
Db 547 YQLPFTTLTSGAVFDMGDAANFKRMKLGSGLVGRWFSPLAPFSDIAYGHSKKIR 606
QY 601 WHISLGRF 609
Db 607 WHISLGRF 615
RESULT 4
AAY74861
ID AAY74861 standard; Protein; 615 AA.
XX AAY74861;
XX 21-MAR-2000 (first entry)
DT Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1196.
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-00989994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 98US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tetelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53623.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 667; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
CC represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 615 AA;
Query Match 98.8%; Score 3136; DB 21; Length 615;
Best Local Similarity 98.7%; Pred. No. 9.1e-249;
Matches 601; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 MMKPTALLPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLPKFPVRIQTQ 60
DB 7 MMKPTALLPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLPKFPVLIQTQ 66
QY 61 DSEIKDMVEEHLPLITQQEEVLDKEQTGFLEAEAPDNVKTMLRSKGYFSSKVSLETKDG 120
DB 67 DSEIKDMVEEHLPLITQQEEVLDKEQTGFLEAEAPDNVKTMLRSKGYFSSKVSLETKDG 126
QY 121 AYVHITPGPRTKIANVGVAIIGDILSDGNLAEEYRNALENWQPVGSDFDQDQSWENSKT 180
DB 127 AYVHITPGPRTKIANVGVAIIGDILSDGNLAEEYRNALENWQPVGSDFDQDQSWENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNDPDTAVLNVVVDVSGRPIAFGDFEITGRTQYPEQIV 240
DB 187 SVLGAVTRKAYPLAKLGNTRAAVNDPDTAVLNVVVDVSGRPIAFGDFEITGRTQYPEQIV 246
QY 241 SGLARFQCTPYDILLDDLLDQALPONGHYSGASVQADFRLQGRVVPKVSVTEVKRKH 300
DB 247 SGLARFQCPMPYDILLDDLLDQALPONGHYSGASVQADFRLQGRVVPKVSVTEVKRKH 306
QY 301 LETGIRLSEYGLGCKIAYDYNNLFNKGIGSVVWMDMKYETTTLAAGTSQPRNRYGNWT 360
DB 307 LETGIRLSEYGLGCKIAYDYNNLFNKGIGSVVWMDMKYETTTLAAGTSQPRNRYGNWT 366
QY 361 SNVSYNRSTTONLEKRAFSGGIWIYVRDRAGIDARLGAEFLAEGRKIPGSDIDLGNSHATM 420
DB 367 SNVSYNRSTTONLEKRAFSGGIWIYVRDRAGIDARLGAEFLAEGRKIPGSDIDLGNSHATM 426
QY 421 LTASHKROLLNNVLPENGHYLDGKIGTTLGTLFTLSSTALIRTSARAGYFFTPENKKLCTF 480
DB 427 LTASHKROLLNNVLPENGHYLDGKIGTTLGTLFTLSSTALIRTSARAGYFFTPENKKLCTF 486

QY 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNGSVLPERALLVGSLE 540
DB 487 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNGSVLPERALLVGSLE 546
QY 541 YOLPPTRTLSCGAVFHDMDGAAANFKRMKLKHGSLGCVRWFPPLAPFSPDIAYGHSDKKIR 600
DB 547 YOLPPTRTLSCGAVFHDMDGAAANFKRMKLKHGSLGCVRWFPPLAPFSPDIAYGHSDKKIR 606
QY 601 WHISLGTRF 609
DB 607 WHISLGTRF 615

RESULT 5

AAZ99625

ID AAY99625 standard; Protein: 587 AA.

XX

AC AAY99625;

XX

DT 03-JAN-2001 (first entry)

XX

DE Neisseria meningitidis BASB040 putative protein sequence.

XX BASB040; bacterial disease; respiratory tract infection; bacteraemia;

KW meningitis; cancer; autoimmune disease.

XX

OS Neisseria meningitidis.

XX

PN WO200034480-A1.

XX

PD 15-JUN-2000.

XX

PF 02-DEC-1999; 99WO-EP09560.

XX

PR 07-DEC-1998; 98GB-0026886.

XX

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Ruelle J;

XX

DR WPI; 2000-423426/36.

DR N-PSDB; AAA48509.

XX

PT Novel BASB040 polypeptides of Neisseria meningitidis useful for

PT diagnostic, prophylactic and therapeutic purposes against microbial

PT diseases comprise a specific amino acid sequence -

XX

PS Claim 4; Page 62; 98pp; English.

XX

CC The present sequence is a putative version of the Neisseria
CC meningitidis strain H44/76 BASB040 protein sequence. This protein is
CC similar to the D15 outer membrane protein of the bacterium. The protein is
CC its gene, antibodies, antagonists and agonists can be used to diagnose
CC and treat bacterial diseases such as those leading to upper respiratory
CC tract infections, bacteraemia and meningitis. In addition, they can be
CC used in vaccines for use against cancer and autoimmune diseases.

SQ Sequence 587 AA;

Query Match 95.1%; Score 3019; DB 21; Length 587;

Best Local Similarity 98.6%; Pred. No. 3.5e-239;

Matches 579; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 23 AADLSENKAAGFALFNKSPDTSVKLPKFPVRIQTQDSEIKDMVEEHLPLITQQQEEV 82

DB 1 AADLSENKAAGFALFNKSPDTSVKLPKFPVLIQTQDSEIKDMVEEHLPLITQQQEEV 60

QY 83 LDKBQGTGLAEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVHITPGPRTKIANVGVAI 142

DB 61 LDKBQGTGLAEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVHITPGPRTKIANVGVAI 120

QY 143 GDILSDGNLAEEYRNALENWQPVGSDFDQDQSWENSKTSVLGAVTRKGYPLAKLGNTRAA 202

Db 121 GDLSDGNLAAYRNALNNQPPVSGDFDQDSWENSKTSVLGAVTRKAYPLAKLGNTOAA 180
QY 203 VNPDDTATVDLNVVDSGRPTAFGDFEITGTORYPEQIVSGLARFQPGTVPVDLDDLLDFQ 262
Db 181 VNPDDTATVDLNVVDSGRPTAFGDFEITGTORYPEQIVSGLARFQPGMPYDLDLLDFQ 240
QY 263 ALBQNGHYSASVQADFDRLQGRVPVKVSVTEVKRHKLETGIRLDSYGLGKGIAYDIY 322
Db 241 ALBQNGHYSASVQADFDRLQGRVPVKVSVTEVKRHKLETGIRLDSYGLGKGIAYDIY 300
QY 323 NLFNKYIGSVVWMDKYEITTLAAGISQPNRYGNWTNVSYNRSTTQNLKRAFSGGI 382
Db 301 NLFNKYIGSVVWMDKYEITTLAAGISQPNRYGNWTNVSYNRSTTQNLKRAFSGGV 360
QY 383 WYVDRAGIDARLGAFLAEGRIKPGSDIDLGNSHATMLTASWKROLLNNVLPENGHYL 442
Db 361 WYVDRAGIDARLGAFLAEGRIKPGSAVDLGNSHATMLTASWKROLLNNVLPENGHYL 420
QY 443 DGKIGTTLGTFLLSTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTVARDNADVP SGLM 502
Db 421 DGKIGTTLGTFLLSTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTVARDNADVP SGLM 480
QY 503 FRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLEYQLPTRLTSGAVFHDMDGAAA 562
Db 481 FRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLEYQLPTRLTSGAVFHDMDGAAA 540
QY 563 NFRKMLKHGSGLVGRWFSPLAFSPFDIAGHSDKKIRWHISLGTRF 609
Db 541 NFRKMLKHGSGLVGRWFSPLAFSPFDIAGHSDKKIRWHISLGTRF 587
RESULT 6
AA74860
ID AA74860 standard; Protein; 522 AA.
XX
AC AA74860;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 286 protein sequence SEQ ID NO:1194.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99NO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AA253622.
XX

PT Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
PS Claim 2; Page 666-667; 1453pp; English.
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 522 AA;
Query Match 45.5%; Score 1443; DB 21; Length 522;
Best Local Similarity 91.7%; Pred. No. 9e-110;
Matches 287; Conservative 2; Mismatches 16; Indels 8; Gaps 2;
QY 1 MMKPTALLPALFFPHAYAPAADLSENKAAGFALFKKSPDTSVKLPKFPVRIDTQ 60
Db 7 MMKPTALLPALFFPHAYAPAADLSENKAAGFALFKKSPDTSVKLPKFPVRIDTQ 66
QY 61 DSEIKDMVEEHLPLITQOOEEVLDEKQGTGLAEAPDNVKTMLRSKGYFSKVSLETKDG 120
Db 67 DSEIKDMVEEHLPLITQOOEEVLDEKQGTGLAEAPDNVKTMLRSKGYFSKVSLETKDG 126
QY 121 AYTIVHTPGPRTKIANGVAILGDILSDGNLAEYRNALNNQPPVSGDFDQDSWENSKT 180
Db 127 AYTIVHTPGPRTKIANGVAILGDILSDGNLAEYRNALNNQPPVSGDFDQDSWENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPTAFGDFEITGTORYPEQIV 240
Db 187 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPTAFGDFEITGTORYPEQIV 246
QY 241 SGLARFQPGTYPYDLDLLDFQALQNGHYSGASVQADFRL-----QGDRVPVKVSV 293
Db 247 SGLARFQPGTYPYDLDLLDFQALQNGHYSGASVQADFRLPRGPRPROSRNRGQTQ 306
QY 294 TEVVRHKLKLETGIR 306
Db 307 TR-NRHPPPREFGIR 318
RESULT 7
AA75820
ID AA75820 standard; Protein; 578 AA.
XX
AC AA75820;
XX
DT 07-NOV-2000 (first entry)
XX
DE Haemophilus influenza strain Rd KW20 BASB067 protein.
XX
KW BASB067; outer membrane protein; antigen; vaccine; antibiotic;
KW antibacterial; screening; infection; diagnosis; therapy.
XX
OS Haemophilus influenzae.
XX
FH Key Location/Qualifiers
FT Peptide 1..22 /label= signal_peptide
FT Protein 23..578 /label= Mature_protein
FT Domain 23..236 /note= "N-terminal domain"
FT Domain 237..578

FT XX /note= "C-terminal domain"
PN WO200047737-A1.
XX 17-AUG-2000.
PD 04-FEB-2000; 2000WO-EP00887.
XX 09-FEB-1999; 99GB-0002880.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA Ruelle J, Thonnard J;
XX WPI; 2000-515059/46.
XX N-PSDB; AAA50269.
XX BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
PT used for diagnosing and treating H. influenzae infections -
XX Claim 1; Page 78-80; 87pp; English.
XX The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae strain Rd KW20. BASB067 is a surface
CC expressed protein that is recognised by the immune system. It
CC shows 23% identity to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its
CC N-terminal domain is predicted to contain a mixture of alpha-helix
CC and beta-strands, and could be used as a vaccine antigen. The
CC C-terminal domain is predicted to form a beta-barrel composed of
CC anti-parallel, amphipathic beta-strands. The external loops of
CC the beta-barrels of integral outer membrane proteins frequently
CC contain immunodominant B-cell epitopes, making the C-terminal
CC domain of BASB067 a strong candidate vaccine antigen. The
CC invention relates to recombinant materials and methods for the
CC production of BASB067 polypeptides and polynucleotides, for use
CC especially in therapeutic and prophylactic vaccines. It also
CC relates to methods for using such polypeptides and polynucleotides
CC in the prevention and treatment of microbial diseases, in diagnostic
CC assays for detecting diseases associated with microbial infections,
CC and assays for detecting expression or activity of BASB067
CC polypeptides or polynucleotides. Antibodies raised against
CC BASB067 can be used to treat humans with H. influenzae disease.
XX
SQ Sequence 578 AA;
Query Match 11.8%; Score 374; DB 21; Length 578;
Best Local Similarity 22.9%; Pred. No. 7.1e-22;
Matches 153; Conservative 93; Mismatches 263; Indels 158; Gaps 18;
QY 3 IKPTALLPALFFPPHAYAPADLSENKAAGFALFKNKSPDTSVKLKPFPVRIQTQDS 62
DB 6 LKLTALFL-ALSCPPAPAEQTVDIEVQIRGFRVNTD-----LNVLINK 51
QY 63 EIKDNVEHLPLITQOQEEVLDEKQTFLEAEADPNVKTMLRSKYFSSKVSLETKD--- 119
DB 52 EEMDGSERYQHLVTKA-----VDRLRVFGYVESSVFRPKRQOG 91
QY 120 --GAYTVHITPGPKTKANYGVATILGDILSDGNLAERYRNALNNKQOPVGSDFDQDSWEN 177
DB 92 KRDLIIAHVTPGETKTAGTDVQTEGGAQDENF-----NALRKNLPKDGVLVVEHQTYDD 146
QY 178 SKTSVLGAVTRKGYPLAKLGNRAAVNPDTATVDLNVVDSGRPIAFGDFEITQRTYPE 237
DB 147 YKTAISRLALNRGYFDGNEFKISRLSEIPETHQAWRMFLDSGVRYHYGNITFSSHQ-IRD 205
QY 238 QIVSGLARFQPGTGYDLDLLDFQOALEQNGHYSGASVQADFRLQGDGRVPVKVSVTEVK 297
DB 206 DYLNILNIKSGDPYLMNNLSDLTSDFPSSNWFSSVLVQPNVNH-KSKTVDVVEILIYPRK 264
QY 298 RHKLETGIRLDSVGLGCKTAY----- 319
DB 265 KNAMELVGFGSDGVGHQIGWTKPWINSRHSLSRNLNLSAPKQTLNRYMPLLNPL 324

QY 320 DYNNLFNKGYIGSVVWDMKYETTTAAAGISQPRNRYRGNWYTS-----NVSYNRST 369
DB 325 NYYVDFAVGVEGEKENDTNRVLTLTA-----LRYWNNAHGWQYEGGLRNRYSFT 375
QY 370 TONLEKRAF-----SGGIWYVDRAGIDARLCAEFLAGRKIPGSDIDLGNSHATMLTASW 425
DB 376 QADITDKTLLLYPTVGTTRRLRGGSFATWG-----DVQKI---TFDLS----- 416
QY 426 KROLLNNVLHPENGHYLDGKIGTITGLTFLSTALIRTSARAGYFTT-PENKLGKLTFTIIRG 484
DB 417 KR-----IWLSESEPIKVOASSAWVRYTAENHRV---VARA 449
QY 485 QAGYTVARNADVPVSGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLP 544
DB 450 EIGYLHTKGIEKIPPTLRFAGGDRSVRGYKKIAPKNNRNGKLVGSRLLTTLSLEYQYQ 509
QY 545 FTRTSLGAVHDMGDAANFKRMKLGKHSGLGVRFWFLAPFSDIAYGHS-----KKIR 600
DB 510 VYPNWAATFADSGLAADNVTAKELRYCTGVGRWASPVGAIKFPIATPIINDKNSKNIQ 569
QY 601 WHISLGT 607
DB 570 FYIGLGT 576
RESULT 8
ID AAY95821 standard; Protein; 576 AA.
XX AAY95821;
XX 07-NOV-2000 (first entry)
XX Haemophilus influenza nH strain 289 BASB067 protein.
DE BASB067; outer membrane protein; antigen; vaccine; antituberc;
KW antibacterial; screening; infection; diagnosis; therapy.
XX Haemophilus influenzae.
FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..576
FT /label= Signal_peptide
FT /label= Mature_protein
XX WO200047737-A1.
XX 17-AUG-2000.
XX 04-FEB-2000; 2000WO-EP00887.
XX 09-FEB-1999; 99GB-0002880.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Ruelle J, Thonnard J;
XX WPI; 2000-515059/46.
XX N-PSDB; AAA50270.
XX BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
PT used for diagnosing and treating H. influenzae infections -
XX Claim 1; Page 81-82; 87pp; English.
XX The present sequence is that of outer membrane protein BASB067 of
CC Haemophilus influenzae non-typeable (NHI) strain 289. BASB067 is
CC a surface expressed protein that is recognised by the immune system.
CC It shows homology to the protective surface antigen D15 of
CC H. influenzae and has a similar secondary structure. Its
CC N-terminal domain is predicted to contain a mixture of alpha-helix

Db 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNKAPKRYADFQK 554
 QY 403 GRKIPGSDIDLGNHATML---TASWKROLLNNVLHPENGHYLDGKIGTTLGTLSLSTAL 459
 Db 555 YKRTDGDAD---GSFKGLLYKGTVCWGRNKTDLSALWTRG-YLTG-----VNAEI 599
 QY 460 IRTSARAGYFTTPNK-----KLGTFTIRGQAGYTVARDNA-DVPSGLMFRSGGASSV 511
 Db 600 ALPGSKLQYYSATHNQTFWFPPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENFYGGGLGSV 659
 QY 512 RGYELDSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTSLGAVFHDMD 559
 Db 660 RGYE---SGTLGPKVYDEYGEKISYGNKKANVASAELLFPMPCAKDARTVRLSLFADAGS 716
 QY 560 -----AAANFKRM-----KLKHGSGLGVRWFSPAPFSDIAY---G 593
 Db 717 VWDGRTYTAENGNNKSVSYSENAHKSTFTNELRYSAGGAVTWLSPLGPMKFIYAPLKKK 776
 QY 594 HSDKKIRWHISLGRF 609
 Db 777 PEDEIQRFQFOLGTF 792

RESULT 10
 AAU03958
 ID AAU03958 standard; Protein; 792 AA.
 AC AAU03958;
 DT 23-OCT-2001 (first entry)
 DE Neisseria meningitidis serogroup A antigenic protein #1.
 KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
 KW bacterial infection; baculovirus; yeast.
 OS Neisseria meningitidis.
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "Signal peptide"
 FT Protein 22..792
 FT /note= "Mature N. meningitidis serogroup A antigen"
 PN WO200138350-A2.
 PD 31-MAY-2001.
 PP 28-NOV-2000; 2000WO-IB01851.
 PR 29-NOV-1999; 99GB-0028197.
 PR 09-MAR-2000; 2000GB-0005698.
 PA (CHIR-) CHIRON SPA.
 PA (STAT-) STATENS INST FOLKEHELSE.
 PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
 DR WPI; 2001-381289/40.
 DR N-PSDB; AAS07278.
 PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
 PT gonorrhoeae, useful in the manufacture of a medicament for treating and
 PT preventing Neisserial bacteria infection -
 XX Claim 1; Page 66-68; 92pp; English.
 CC The sequence represents a Neisseria meningitidis serogroup A 85 kDa
 CC antigenic protein. Neisseria meningitidis colonises the pharynx, causing
 CC meningitis and, occasionally, septicaemia in the absence of meningitis.
 CC This antigenic protein is useful in the manufacture of a medicament for
 CC treating or preventing infection due to Neisseria bacteria, such as
 CC meningitis and septicaemia. It is also useful as a diagnostic reagent for

CC detecting the presence of Neisseria bacteria or antibodies raised against
 CC Neisseria, and as a reagent for raising the antibodies. The Neisserial
 CC nucleotide sequences can be expressed in a variety of different
 CC expression systems, for example, mammalian cells, baculoviruses, plants,
 CC bacteria and yeast.
 CC Note: There are two versions of this sequence displayed in the
 CC specification (see AAU04451).
 XX
 SQ Sequence 792 AA;

Query Match 8.1%; Score 256.5; DB 22; Length 792;
 Best Local Similarity 22.4%; Pred. No. 5.1e-12;
 Matches 138; Conservative 86; Mismatches 243; Indels 149; Gaps 28;

QY 97 DNVTMLRSKGYF-----SSKVSJTEKDGAYTVHIT--PGPRTKIANGVAILGDILSDG 149
 Db 223 EKVYDFYONNGYDFRILDDTDIQTNEKTKTQTIKIVHEGGRFMGKVS-----EG 274
 QY 150 NLAEEYRNALE-----NQOQPVGSDFDQDSWENKTSVLGAVTRK---GYPLAKLG 197
 Db 275 DTNEVPKRAELEKLLTMKPKWYE-----RQOFTAVLGEIONRMGSAGYAYSEI- 322
 QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFGDFEITQRYPEOIVSGLARFQPTPYDLDLL 257
 Db 323 SVQPLPNAGTKTVDFVLHIEPGRKIYVNEIHTGNKTRDEVVRRELQRQMESAPYDTSKL 382
 QY 258 LDFQOALEONCHYSGASVQADFRLQG--DRVVPKVSVTEVKRHKLETGIRLDSFYGLGG 315
 Db 383 QRSKERVLLGYFD--NVQFDVAVPLAGTPDKVDLNNLSLTERSTGSLDLSAGWVODTGLVM 440
 QY 316 KIAYDYNNLFNKYIGSVVMDMKYETTLAAGISQPRNYRGNYWTS---NVSYN----- 366
 Db 441 SAGYSQDNFLGTG--KSNALRASRSKTTLNGSLs---FTDPYFTADGVSLGYDIYKAP 494
 QY 367 -----RSTQNLKRAFSGG-----IWYVRDRAGIDA-----RLGAEFLAE 402
 Db 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNKAPKRYADFQK 554
 QY 403 GRKIPGSDIDLGNHATML---TASWKROLLNNVLHPENGHYLDGKIGTTLGTLSLSTAL 459
 Db 555 YKRTDGDAD---GSFKGLLYKGTVCWGRNKTDLSALWTRG-YLTG-----VNAEI 599
 QY 460 IRTSARAGYFTTPNK-----KLGTFTIRGQAGYTVARDNA-DVPSGLMFRSGGASSV 511
 Db 600 ALPGSKLQYYSATHNQTFWFPPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENFYGGGLGSV 659
 QY 512 RGYELDSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTSLGAVFHDMD 559
 Db 660 RGYE---SGTLGPKVYDEYGEKISYGNKKANVASAELLFPMPCAKDARTVRLSLFADAGS 716
 QY 560 -----AAANFKRM-----KLKHGSGLGVRWFSPAPFSDIAY---G 593
 Db 717 VWDGRTYTAENGNNKSVSYSENAHKSTFTNELRYSAGGAVTWLSPLGPMKFIYAPLKKK 776
 QY 594 HSDKKIRWHISLGRF 609
 Db 777 PEDEIQRFQFOLGTF 792

RESULT 11
 AAB84745
 ID AAB84745 standard; Protein; 792 AA.
 XX
 AC AAB84745;
 XX
 DT 17-SEP-2001 (first entry)
 DE Amino acid sequence of a Neisseria gonorrhoeae protein.
 XX Serogroup B protein; outer membrane protein; Neisserial infection;
 KW vaccine.
 XX
 OS Neisseria gonorrhoeae.

QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTLGAVFHDMDG----- 559
ID AAU03959 standard; Protein; 797 AA.
XX
AC AAU03959;
DT 23-OCT-2001 (first entry)
DE Neisseria gonorrhoeae antigenic protein.
XX Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.
XX Neisseria gonorrhoeae.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note= "Signal peptide"
FT Protein 22..797
FT Protein /note= "Mature N. gonorrhoeae antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PE 28-NOV-2000; 2000WO-1B01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX
DR WPI; 2001-381289/40.
DR N-PSDB; AAS07279.
XX
PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing Neisserial bacteria infection -
XX
PS Claim 1; Page 37-39; 92pp; English.
XX
CC The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.
CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the
CC pharynx, causing meningitis and, occasionally, septicaemia in the absence
CC of meningitis. This antigenic protein is useful in the manufacture of a
CC medicament for treating or preventing infection due to Neisseria
CC bacteria, such as meningitis and septicaemia. It is also useful as a
CC diagnostic reagent for detecting the presence of Neisseria bacteria or
CC antibodies raised against Neisseria, and as a reagent for raising the
CC antibodies. The Neisserial nucleotide sequences can be expressed in a
CC variety of different expression systems, for example, mammalian cells,
CC baculoviruses, plants, bacteria and yeast.
XX
SQ Sequence 797 AA;

Query Match 8.0%; Score 253; DB 22; Length 797;

Best Local Similarity 21.9%; Pred. No. le-11;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;
QY 97 DNVTMLRSKGYF-----SSKVSLETKDGAYTVHIT--PCPRTKIANVGVAIGLDILSDG 149
DB 223 EKVTDFYNNGYDFRILDFDIQTNEKTKQTIKITVHEGGRFRWGVKVS-----EG 274
QY 150 NLAEYRNAL-----NWQPVGSDFDODSWENSKTSVLGAVTRK-----GYPLAKLG 197
DB 275 DTNEVPKAELEKLLTMKPGKWE-----ROOMTAVLGEIQNRWGSAGYAYSEI- 322
QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFQDFEITGTORYPEQIVSGIARPOGTPYDLDLL 257
DB 323 SVQPLPNAETKTVDVFLHIEPGRKIYVNEIHITGNKTRDEVVRRRELROMESAPYDTSKL 382
QY 258 LDFQOALEQNGHYSAGSQADFRLQ--DRVPVKYSVTEVKRHKLETGIRLSEVGLGG 315
DB 383 QRSKREVELGYFD--NVQFDVPLAGTPDKVDLNMSLTERSTGSLDLSAGWYQDTGLVM 440
QY 316 KIAYDYNLFNKGYIGSVVMDMDKYETL-----AAGISQPRNYRGNW----- 359
DB 441 SAGVSODNLEFGT--KSAALRASRSKTTLNGSLSTDPYFTADGVSLGYDYVYKAFDPRK 498
QY 360 -TSNVSYNRSTTONLEKRAFSGGIWYVRDAGIDA-----RLGAELAEGRKI 406
DB 499 ASTSIQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHYADFTKYGKT 558
QY 407 PGSDIDLGNHATML--TASWKRQLLNNVLPENGHYLDGKIGTTLGTLFSLTALRTSA 464
DB 559 DGTD---GSFKWLKYGTVGWRNKDTSALWPTRG-YLTG-----VNAEIALPGS 604
QY 465 RAGYFTFPENK-----KLGTFTIRGOAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
DB 605 KLQYYSATHNQTFWFFPLSKTFTLMLGGEVGIAGYGTKEIPFENFYGGGLGSVRGYE- 663
QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTLGAVFHDMDG----- 559
DB 664 --SGTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLFADAGSVMDGK 721
QY 560 -----AAANFKRM-----KLKHGSLGVWVWFSPLAPFFSDTAY-----G 593
DB 722 TYDDNSSATGGRVQNIYAGNTHKSTFTNELRYSAGAVTWLSPLGPMKFSYAYPLKKK 781
QY 594 HSDKKIRWHISLGTGF 609
DB 782 PEDEIORFQQLGTTF 797
RESULT 15
AAU04451
ID AAU04451 standard; Protein; 797 AA.
XX
AC AAU04451;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein #2.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note= "Signal peptide"
FT Protein 22..797
FT Protein /note= "Mature N. meningitidis serogroup A antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX


```
QY 97 DNKTMRLSRKGYF-----SSKVSUTEKDGAYTVHIT--PGPRTKIANVGVAILGDILSDG 149
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 EKVTDYFQNGYDFRILDTDIQTNEKTKQITIKITVHEGGRFRWKGKVI-----EG 274
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 150 NLAEYYRNALE-----NQQPVGSDFDQDQSWENSKTSVLGAVTRK-----GYPLAKLG 197
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 DTNEVPKRAELEKLLTMKPGKWE-----RQOMTAVLGEIQNRMGSAAYSEI- 322
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFDPEITGTORYPEQIVSGLARFQPGTYPDLDLL 257
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 SVQPLPNAETKTVDVLHIEPGRKIYVNEIHITGNKTRDEVVRELRQMESAPYDTSKL 382
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 258 LDFQALQNGHYSGASVQADFRLQ--DRVPVKVSVTEVKRHKLETGIRLDSYGLGG 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 QRSKERVLLGYFD--NVQFDVAVPLAGTPDKVDLNLMSLTERSTGSLDLSAGWQDTGLVM 440
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 316 KIAVDYNNLFNKGYIGSVVMDMKYETTL-----AAGISQPNRYGNW----- 359
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 SAGVSQDNLFCTG--KSAALRASRSKTTLNGSLSFDPYFTADGVSLGYDVYKAFDPRK 498
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 360 -TSNVSYNRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAELAEGRKI 406
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 499 ASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHADFIKKYGKT 558
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 407 PGSDIDLGNSHATML--TASWKROLLNNVLHPENGHYLDGKIGTTLGTFLSSTALIRTS 464
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 DGTD---GSFKGWLYKCTVWGGRNKTDLSALWPTRG-YLTG-----VNAEIALPGS 604
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 465 RAGYFFTPENK-----KLGTFTIIRQAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 KLOYSATHNQTFWFFPLSKTFTLMLGGEVGTAGGYGRTKEIPFPENFYGGGLGSRGYE- 663
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 517 DSIGLAGPN-----GSVLP-----BRALLVGSLEYQLP---FTRTLGSAVFHDMGD----- 559
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 --SGLTGPVKYVDEYGEKISYGNKKANVSALLFPMPGAKDARTVRLSLFADAGSVMDGK 721
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 560 -----AAANFKRM-----KLHGSGLGVRWFSPPLAFPSFDIAY-----G 593
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 TYDDNSSATGGRVQNIYAGNTHKSTFTNELRYSAGGAVTWLSPLGPMKFSYAYPLKKK 781
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 594 HSDKKIRWHISLGRF 609
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 782 PEDEIQRFQQLGTTF 797
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 17
AAB84746
ID AAB84746 standard; Protein: 797 AA.
AC
XX
XX AAB84746;
XX
XX
XX 17-SEP-2001 (first entry)
XX
XX Amino acid sequence of a Neisseria serogroup A protein.
XX
XX Serogroup A protein; outer membrane protein; Neisserial infection;
XX
XX vaccine.
XX
XX Neisseria meningitidis.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX
XX WO200152885-A1.
XX
XX
XX 26-JUL-2001.
XX
XX 17-JAN-2001: 2001WO-IB00166.
XX
XX
```

```
PR 17-JAN-2000: 2000GB-0001067.
XX 09-MAR-2000: 2000GB-0005699.
PA (CHIR-) CHIRON SPA.
XX
XX Pizza M, Rappuoli R, Giuliani M;
PI
XX WPI: 2001-451895/48.
XX N-PSDB: AAH42130.
XX
XX Composition for treating or preventing infection to, detecting, or for
XX raising antibodies against Neisserial bacteria, comprises an N.
XX meningitidis serogroup B outer membrane preparation and an immunogenic
XX component -
XX
XX Disclosure: Page 71-74; 83pp; English.
XX
XX The present sequence represents a Neisseria serogroup A protein. The
XX protein is used to produce the compositions of the invention. The
XX specification describes a composition, comprising a Neisseria
XX meningitidis serogroup B outer membrane preparation and an immunogenic
XX component. The immunogenic component is protein disclosed in WO99/57280,
XX WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
XX WO99/31132, WO99/38683, WO99/55873, and/or N. meningitidis protein PorA,
XX TbpA, TbpB, PilC, OpA, or Omp85. The composition is used for making
XX a medicament for treating or preventing infection due to Neisserial
XX bacteria; a diagnostic reagent for detecting the presence of Neisserial
XX bacteria or of antibodies raised against Neisserial bacteria; and/or
XX a reagent which can raise antibodies against Neisserial bacteria. It may
XX also be used as a vaccine.
XX
XX Sequence 797 AA;
XX
XX Query Match 8.0%; Score 253; DB 22; Length 797;
XX Best Local Similarity 21.9%; Pred. No. le-11;
XX Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;
XX
XX QY 97 DNKTMRLSRKGYF-----SSKVSUTEKDGAYTVHIT--PGPRTKIANVGVAILGDILSDG 149
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 EKVTDYFQNGYDFRILDTDIQTNEKTKQITIKITVHEGGRFRWKGKVI-----EG 274
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 150 NLAEYYRNALE-----NQQPVGSDFDQDQSWENSKTSVLGAVTRK-----GYPLAKLG 197
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 DTNEVPKRAELEKLLTMKPGKWE-----RQOMTAVLGEIQNRMGSAAYSEI- 322
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFDPEITGTORYPEQIVSGLARFQPGTYPDLDLL 257
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 SVQPLPNAETKTVDVLHIEPGRKIYVNEIHITGNKTRDEVVRELRQMESAPYDTSKL 382
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 258 LDFQALQNGHYSGASVQADFRLQ--DRVPVKVSVTEVKRHKLETGIRLDSYGLGG 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 QRSKERVLLGYFD--NVQFDVAVPLAGTPDKVDLNLMSLTERSTGSLDLSAGWQDTGLVM 440
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 316 KIAVDYNNLFNKGYIGSVVMDMKYETTL-----AAGISQPNRYGNW----- 359
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 SAGVSQDNLFCTG--KSAALRASRSKTTLNGSLSFDPYFTADGVSLGYDVYKAFDPRK 498
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 360 -TSNVSYNRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAELAEGRKI 406
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 499 ASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHADFIKKYGKT 558
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 407 PGSDIDLGNSHATML--TASWKROLLNNVLHPENGHYLDGKIGTTLGTFLSSTALIRTS 464
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 DGTD---GSFKGWLYKCTVWGGRNKTDLSALWPTRG-YLTG-----VNAEIALPGS 604
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 465 RAGYFFTPENK-----KLGTFTIIRQAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 KLOYSATHNQTFWFFPLSKTFTLMLGGEVGTAGGYGRTKEIPFPENFYGGGLGSRGYE- 663
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 517 DSIGLAGPN-----GSVLP-----BRALLVGSLEYQLP---FTRTLGSAVFHDMGD----- 559
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 --SGLTGPVKYVDEYGEKISYGNKKANVSALLFPMPGAKDARTVRLSLFADAGSVMDGK 721
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 560 -----AAANFKRM-----KLHGSGLGVRWFSPPLAFPSFDIAY-----G 593
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 TYDDNSSATGGRVQNIYAGNTHKSTFTNELRYSAGGAVTWLSPLGPMKFSYAYPLKKK 781
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 594 HSDKKIRWHISLGRF 609
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 782 PEDEIQRFQQLGTTF 797
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 17
AAB84746
ID AAB84746 standard; Protein: 797 AA.
AC
XX
XX AAB84746;
XX
XX
XX 17-SEP-2001 (first entry)
XX
XX Amino acid sequence of a Neisseria serogroup A protein.
XX
XX Serogroup A protein; outer membrane protein; Neisserial infection;
XX
XX vaccine.
XX
XX Neisseria meningitidis.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein /note= "signal peptide"
XX /note= "mature protein"
XX
XX WO200152885-A1.
XX
XX
XX 26-JUL-2001.
XX
XX 17-JAN-2001: 2001WO-IB00166.
XX
XX
```


XX

PS Claim 22; Page 33; 39pp; English.

XX The present invention describes an immunogenic composition (I) comprising a Neisseria antigen and an adjuvant composition comprising an oligonucleotide comprising at least 1 CG motif. Also described is an adjuvant composition (II) comprising an oligonucleotide which comprises at least 1 CG motif and a complete Freund's adjuvant (CFA), where the oligonucleotide preferably comprises at least one phosphorothioate bond. AA92359 to AA92385 represent specifically claimed oligonucleotides of the present invention. (I) is useful for stimulating an immune response in a mammal, preferably a human, against Neisseria infection, preferably Neisseria meningitidis infection and in the manufacture of a medicament for inducing a protective immune response in a mammal. The present CC sequence represents the claimed Neisseria meningitidis serogroup A amino CC acid sequence disclosed in GB-9928197.4, which is given in the present CC invention.

XX Sequence 797 AA;

Query Match 8.0%; Score 253; DB 22; Length 797;

Best Local Similarity 21.9%; Pred. No. 1e-11;

Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

QY 97 DNVTMLRSKGYF-----SSKVSUTEKDGYATVHIT--PGPRTKIANVGVAILEDLSDG 149

Db 223 EKVTDFYQNNGYDFRILDTIQTNEDKTKQTKITVHEGGRFRWGVSI-----EG 274

QY 150 NLAEYYRNALE-----NMQQPVGSDFDQDSDWENSKTSVLGAVTRK-----GYPLAKLG 197

Db 275 DTNEVPAEAELEKLLTMKPGKWE-----ROMTAVLGEIQNRMSAGYAYSEI- 322

QY 198 NTRAAVNPDTATVNLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFOPGTPYDLDLL 257

Db 323 SVQPLPNAETKTVDVFLHIEPGRKIYVNEIHTGNNTKTRDEVVRRELQWESAPYDYSKL 382

QY 258 LDFQOALEQNGHYSVASQVADFRLQ--DRVPVKVSVTEVKRHKLTGIRLSEYGLGG 315

Db 383 QRSKERVELLGYFD--NVQFDAVPLAGTPDKVDLNMSLTERSTGSLDSAGWQDTGLVM 440

QY 316 KIAVDYNNLFNKGKYGISVWMDKYEYTL-----AAGISQPRNRYGNW---- 359

Db 441 SAGVSQDNLFQGTG--KSAALRASRKTTLNGSLSFDPYFTADGVSILGYDYGKAFDPRK 498

QY 360 -TSNVSYNRSRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAFLAEGRKI 406

Db 499 ASTSIKQYKTTAGAGIRMSVPYTEYDRVNFGLVAEHLTVNTYNKAPKHYADFIKKYGKT 558

QY 407 PGSDIDLGNSHATML--TASWKROLLNNVLPENGHYLDCKIGTTLGTLFSLSTALIRTS 464

Db 559 DGTD--GSFKGMLYKGTGVGRNKTDSALWPTRG--YLTG-----VNABIALPGS 604

QY 465 RAGYFFTPENK-----KLGTFIIRQAGYTVARDNA-DVPSGLMFRSGASSVRGYEL 516

Db 605 KLOQYSATHNQWTFPLSKTFTLMLGGEVGTAGGYGRKTEIPFPENFYGGGLSGVRGYE- 663

QY 517 DSIGLAGPN-----GSLVLP--ERALLVGSLEYOLP--FTRTLGSAVFHDMGD----- 559

Db 664 --SGTLGPKVYDEYGEKISYGGNKKANVSARELLFPMPGAKDARTVRLSFADAGSVWDGK 721

QY 560 -----AAANFKRM-----KLKHGSLGLVGRWFSPLAFPFSDIAY-----G 593

Db 722 TYDSSSSATGGRVQNIYVAGNTHKSTFTNELRYSAGGAVTWLSPLGPMKFSYAYPLKKK 781

QY 594 HSDKKIRWHISLGRFP 609

Db 782 PEDEIQRFQFQLGTF 797

RESULT 20

AA984947

ID AA984947 standard; Protein; 797 AA.

XX

AC

XX AAY84947;

DT 21-AUG-2000 (first entry)

XX Amino acid sequence of outer membrane protein (omp) 85.

DE Outer membrane protein; omp85; gonococcal infection; symptomatic disease;

XX meningococcal infection; protective immune response; vaccine.

KW Neisseria meningitidis.

OS WO200023595-A1.

XX 27-APR-2000.

XX 22-OCT-1998; 98WO-US22352.

XX 22-OCT-1998; 98WO-US22352.

XX (UYMO-) UNIV MONTANA.

XX Judd RC, Manning SD;

XX WPI; 2000-339694/29.

DR N-PSDB; AAA15156.

XX New isolated outer membrane protein 85 of Neisseria gonorrhoeae and N.

PT meningitidis useful for vaccine, therapeutic and diagnostic

PT compositions for gonococcal or meningococcal infections -

XX Claim 41; Page 89-92; 98pp; English.

XX The present sequence represents an outer membrane protein (omp) 85 of

CC Neisseria meningitidis. The omp polypeptides and polynucleotides are

CC useful in compositions for use in the prevention, treatment and diagnosis

CC of non-symptomatic gonococcal infection or meningococcal infection and

CC hybridisation complexes. They are also useful for the detection of

CC also provide diagnostic, therapeutic and antibodies specific omp proteins

CC the treatment or prevention of the infections described above. The

CC humans or animals with N. gonorrhoeae, N. meningitidis, or other

CC Neisseria species. The proteins, antibodies and polynucleotide

CC sequences of the present invention may also be used in the screening

CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 797 AA;

QY 97 DNVTMLRSKGYF-----SSKVSUTEKDGYATVHIT--PGPRTKIANVGVAILEDLSDG 149

Db 223 EKVTDFYQNNGYDFRILDTIQTNEDKTKQTKITVHEGGRFRWGVSI-----EG 274

QY 150 NLAEYYRNALE-----NMQQPVGSDFDQDSDWENSKTSVLGAVTRK-----GYPLAKLG 197

Db 275 DTNEVPAEAELEKLLTMKPGKWE-----ROMTAVLGEIQNRMSAGYAYSEI- 322

QY 198 NTRAAVNPDTATVNLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFOPGTPYDLDLL 257

Db 323 SVQPLPNAETKTVDVFLHIEPGRKIYVNEIHTGNNTKTRDEVVRRELQWESAPYDYSKL 382

QY 258 LDFQOALEQNGHYSVASQVADFRLQ--DRVPVKVSVTEVKRHKLTGIRLSEYGLGG 315

Db 383 QRSKERVELLGYFD--NVQFDAVPLAGTPDKVDLNMSLTERSTGSLDSAGWQDTGLVM 440

QY 316 KIAVDYNNLFNKGKYGISVWMDKYEYTL-----AAGISQPRNRYGNW---- 359

Db 441 SAGVSQDNLFQGTG--KSAALRASRKTTLNGSLSFDPYFTADGVSILGYDYGKAFDPRK 498

QY 360 -TSNVSYNRSRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAFLAEGRKI 406

Query Match 7.9%; Score 251; DB 21; Length 797;

Best Local Similarity 21.9%; Pred. No. 1.5e-11;

Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

```
Db 499 ASTSIKQKTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVNTYNKAPKHYADFICKYKT 558
Qy 407 PGSDIDLCNHSATML--TASWKROLLNNVLPENGHYLDGKIGTTLGTFLLSSTALIRTS 464
Db 559 DGTD---GSPGWLKYGTVGGRNKTDALNPTKG-YLTG-----VNAEIALPGS 604
Qy 465 RAGYFFTPENK-----KLGTFIRGQAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
Db 605 KQOYVSATHNQWTFPLSKTFTLLGGEVGIAGGYGRKTEIPFPENFYGGGLGVRGYE- 663
Qy 517 DSIAGLGN-----GSVLP-----ERALLVGSLEYQLP----PTRLGSAVFHDMGD----- 559
Db 664 --SGLTGPVKVYDEYGEKISYCGNKKANVSABELLPMPGAKDARTVRLSLFADAGSVMDGK 721
Qy 560 -----AAANFKRM-----KLKHGSGLGVRWFSPLAPESFDIAY----G 593
Db 722 TYDNSSSATGGRVONIYGAGNTHKSTFTNELRYSGAGVTLWPLGPMKPRYAYPLKKK 781
Qy 594 HSDKKIRWHISLGRF 609
Db 782 PEDEIQRFQFGTTF 797

RESULT 21
AAG78603
ID AAG78603 standard; Protein; 896 AA.
AC AAG78603;
XX
XX
XX 20-NOV-2001 (first entry)
DT
DE Lawsonia intracellularis protein SEQ ID NO: 5.
KW HtrA; PonA; Hyc; YefW; ABC1; Omp100; Lawsonia intracellularis infection;
KW vaccine.
XX Lawsonia intracellularis.
XX JP2001169787-A.
XX
XX 26-JUN-2001.
XX
XX 20-OCT-2000; 2000JP-0320736.
XX
XX 22-OCT-1999; 99US-0160922.
XX (PFIZ ) PFIZER PROD INC.
XX
XX WPI; 2001-592540/67.
XX
XX Lawsonia intracellularis polynucleotide and encoded protein, used to
PT prevent Lawsonia intracellularis infection -
XX
XX Claim 12; Page 43-45; 67pp; Japanese.
XX
XX The present invention provides isolated polynucleotides encoding HtrA,
CC PonA, Hyc, YefW, ABC1 or Omp100 protein of Lawsonia
CC intracellularis. The sequences can be used in vaccines for the prevention
CC of Lawsonia intracellularis infection. The present sequence is a protein
CC of the invention.
XX
XX SQ Sequence 896 AA;
XX
XX Query Match 7.8%; Score 247; DB 22; Length 896;
XX Best Local Similarity 22.8%; Pred. No. 3.7e-11;
XX Matches 140; Conservative 95; Mismatches 276; Indels 104; Gaps 25;

Qy 64 IKDVEEHLPLI---TQQQEEVLDEQ-----TGFLEAEAPN----VKTMLRSKGYF 109
Db 317 IKDVRIEGLETIKAKTLKKELALTFERNFLSWFTGTGVLREYLERDSTASAYAMNHGV 376
Qy 110 -----SSKVSLETEKGATVHTPGPRTKIANWGCVAILGDILSDGNLAEEYRNAL----- 159
```

```
Db 377 DIQVASEVTENEIGIVITFRVKEGRYKI-----GKIDFGDLIETNEQLLKVTKI 428
Qy 160 ---ENMOOPCVSGDFDQDSWENSKT SVLGAVTRKCYPLAKLGNTRAAVNPDPATVDLNVVV 216
Db 429 DDHKNYBOYFSLVSMQDDVK-----ALTDYSDYGYAFAEV-DLETTKNEEDATIDVTFLL 483
Qy 217 DSGRPIAFAGDFEITGTORYPEQIVSGGLARFQPGPYDLDDLDDFQQALEQNGHYSGASVQ 276
Db 484 DKKQKVFLLRIIVGENTFRDNLRELRLADGDLFNCQHLRRSNECLNRLGYFN----Q 539
Qy 277 ADFRL---QGRVPVVKVSTEVKHKLETGJ---RLDSEYGLGKIAVDYVNLNPKGYIG 331
Db 540 VDTDTLPTGKDDEVDLLVKQGEARTGATGGVGYSTHKKFCVSGSIS--ERNLWKGVIL 597
Qy 332 SVVWDMKYEITLAAGLSOPRNYKGNWTSNVSYN-RSTTONLEKRAFSGGIWYVRORAG 390
Db 598 SIEGFISKSSSLDLSFTNPRVYDTDFGFSNNIYTLRDEWDDFRKKTGY-----DTIR 650
Qy 391 IDARLGA-EFLAEGRKI-----PGSDIDL-GNSHATMLTASWKROLLNNVLP 436
Db 651 LFHPIGEYSSIFVGYRIDQYRLYDIPSTAPRSYLDYQGNKISSVVGTFDSDSRERP 710
Qy 437 ENGH-----YLDGKIGTTLGTFLSSTALIRTSARAGYFTPENKKLGFILIRGQAGYT 489
Db 711 SKGHIAKLLIVEYGGGLG-----GNDNFPKPIAELQGFYSISKKNHIIHWTRAGAA 763
Qy 490 VARDNADVPGLMFRSGGASSVRGYELDSICLAGPN-GSVLPERALLVGSLEYQLPFRRT 548
Db 764 YKSKKVPVPDFRFFIGGIDIRGYDTEDLAPKDPFGEIGGDRMAFLNLEYIWTFOPE 823
Qy 549 LSGAV--FHDMDAAANFKR----MKLKHGSGLGVRWFSPLAPESFDIAYG-----HSD 596
Db 824 LGLALVFFYDIGFOTDSVQTSNPFSLKQSYGLELRWRSPMGDLRF--AYGIPLKNVSG 881
Qy 597 KKI--RWHISLGRF 609
Db 882 KKTRGRFEFSMGQFF 896

RESULT 22
ABG17710
ID ABG17710 standard; Protein; 474 AA.
XX
XX AC ABG17710;
XX
XX 18-FEB-2002 (first entry)
DT
XX Novel human diagnostic protein #17701.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS81897.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
```

PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 20; SEQ ID No 48069; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostic products as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 474 AA;

Query Match 7.0%; Score 222.5; DB 22; Length 474;
 Best Local Similarity 39.5%; Pred. No. 1.5e-09;
 Matches 51; Conservative 19; Mismatches 56; Indels 3; Gaps 1;

QY 480 FIIHQAGYTVARDNADVPGLMPSGSGASSVRCGYELDSICLAGPNSVLPERALVGS 539
 Db 91 FVTGTLGTETGDFDKVPPDPLRSGGASVRCGYELDSICLAGPNSVLPERALVGS 150
 QY 540 EYQLPFTRTLGSAGVFDHMDAAAFKRMKLGSLGVRWFSPFLAFPSFDIAYGHSKK- 598
 Db 151 EYQYNVTKMKGAVFVDSGEAVSDIRSDPKTGTGVGRWESPVGPKLDFAVPVADKDE 210
 QY 599 --IRWHISL 605
 Db 211 HGLQFYIGL 219

RESULT 23

AA95988
 ID AA95988 standard; Protein; 889 AA.

XX AA95988;

XX 05-DEC-2000 (first entry)

XX Moraxella catarrhalis BAS081 mature protein.

XX BAS081; infection; otitis media; pneumonia; sinusitis;
 KW inflammation; therapy; antibacterial; antiinflammatory; vaccine;
 KW diagnosis.

XX Moraxella catarrhalis.

XX Key Location/Qualifiers

FT Misc-difference 335 /note= "Val in sequence of AAY95987"

XX WO200052042-A1.

XX 08-SEP-2000.

XX 23-FEB-2000; 2000WO-EP01468.

XX 26-FEB-1999; 99GB-0004559.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Ruelle J;
 PI
 XX WPI: 2000-587296/55.
 DR N-PSDB; AAA50537.
 XX
 PT New BAS081 polypeptides from Moraxella catarrhalis and polynucleotides
 PT encoding the polypeptides used for treating infections, or as a vaccine
 PT for preventing infections, especially those caused by M. catarrhalis .
 XX
 PS Claim 1; Page 61; 97pp; English.

XX The present sequence is that of the Moraxella catarrhalis strain
 CC Mc2391 (ATCC 43617) BAS081 mature protein. BAS081 is related by
 CC amino acid sequence homology to Neisseria meningitidis omp5 outer
 CC membrane protein. The sequence was deduced from PCR-amplified DNA
 CC (see AAA50537), and shows 99.9% homology to the mature region of a
 CC BAS081 gene translation product (see AAY95987). The invention
 CC provides BAS081 polypeptides, polynucleotides, expression vectors,
 CC host cells, and a process for producing a BAS081 polypeptide. Also
 CC provided are vaccine compositions comprising a BAS081 polypeptide
 CC or polynucleotide, and optionally at least 1 other M. catarrhalis
 CC antigen. A method for diagnosing a M. catarrhalis infection
 CC involves identifying a BAS081 polypeptide, or an antibody that is
 CC immunospecific for it, in a sample. A therapeutic composition
 CC useful in treating M. catarrhalis diseases in humans comprises an
 CC antibody directed against a BAS081 polypeptide. The disease can
 CC be a bacterial infection, e.g. otitis media in infants and children,
 CC pneumonia in elders, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed
 CC speech learning, upper respiratory tract infection, and inflammation
 CC of the middle ear.

XX Sequence 889 AA;

Query Match 6.9%; Score 219; DB 21; Length 889;
 Best Local Similarity 22.0%; Pred. No. 7.3e-09;
 Matches 130; Conservative 74; Mismatches 232; Indels 154; Gaps 18;

QY 42 PDESVKLKPRFPVRIDTQDSEIKDMVEHLPLITQOOEEVLDKEQTGFLAEAPDNVKT 101
 Db 432 PVLETVELTDGILMDISPIEFASNLIQDKLNVAARHLYDMPDORVLAINHDDGVNR 491
 QY 102 MLRSKGFSSKVSLETXGDGATVHTPGPRTKIANVGVAIIGDLSDNLAETRYNALEN 161
 Db 492 SI--LGRISDAVSA-----VARAILPDESENEVIDLPERTALAN 528
 QY 162 WQOPVGSDFDODSWENSKTSVLGAVTRKGYPLAKIGNTRAANPDATVDLNVVYDSGRP 221
 Db 529 RKTPA-----DVTQSKK-----VPLVVFVASKP 552
 QY 222 ----IAFGDFEITGTQRYPEQIVSGLARFQPTPYDLDLLDFOAL-EQNGHYSVASVQ 276
 Db 553 RDGQIGLGWGSDDTGT-----LVTKFEHLNLRDGYQAGAEI- 589
 QY 277 ADFRLOGDRVPVKVSTEVKRRHKLGTGIRLDSEYGLGGKAYDYNFNKGYGSGVWD 336
 Db 590 ----RLSEDKKGVKLYATKPLSHPLNDQLRAT----LG-----YQOEVFGHSTNG 631
 QY 337 MDKYETTLLAAGISQPRNRYGNYWTSNVSYNRSTTONLEKRAFSGGIWVRDRADARLG 396
 Db 632 FDLSTRTRLEHISRTIQNGW---NRTYSRLYRLDKLUKTOAPPETWQDLPVDFVNGKPS 688
 QY 397 AEFLAERKIPGSDIDLGNSHATMLTASWKRQLNNVLRHPENGHYLDGKIGTTLGTFLLS 456
 Db 689 QEALLAGVAV-----HKT-----VADNLVNPARGY-----RQYSLEVGS 723
 QY 457 TALIRTS-----ARAG-----YFTFENKLGTFIIRGQAGYTVARNADVPVSL 501
 Db 724 SGLVSDANMAIARAGISGVSGFNAYGNSRAHQMTGGI----QAGYIWSDFNHPVRL 779


```
PA (CONN-) CONNAUGHT LAB LTD.
XX Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
XX WPI: 1994-200269/24.
DR N-PSDB; AAQ66202.
XX
XX Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
XX Disclosure; Fig. 1E; 161pp; English.
XX
XX Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
XX Sequence 793 AA;
SQ
Query Match 6.7%; Score 213.5; DB 15; Length 793;
Best Local Similarity 21.0%; Pred. No. 1.7e-08;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;
QY 105 SKGYFSKSVLT-----EKDGA-YTVHITPPGPRTK-----IANVG--VAIIGDILSDG 149
: || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 228 NGCYAKAQITKTDVQLNDEKTKVNVITDVNEGLQYDLRSARIIGNLGMSAELEPLLSAL 287
QY 150 NLAEYRNALENKQPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNDP--- 206
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 288 HLNDTFRS-----DIADVENAIKAKLG---ERGY-----GNTTVNSVPDFDD 327
QY 207 -TATVDLNVVVDGRPIAFGDFEITGTQRYPEQIVSGARFQPTPYDLDLDDFOQALE 265
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 328 ANTLAITFVVDAGRRLTVQLRFEGNTVSADSTLRQEMRQEGTWTNSOLVELGKIRLD 387
QY 266 QNGHYSAGVQADFRLQG--DRVPKVSVTEVKRHKLETGIRLDSYGLGKGIADYYN 323
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 388 RTGFFE--TVENRIDPINGSNDEVVYKVKERTGTSINFGIGVGTSGISYQTSIKQDN 445
QY 324 LPNKGIGSVVWDMKYTTTLAAGISQPR-----NTRGNVTSVNSYNRS-TTONLEKR 376
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 446 FLGTGAASVIAGTAKNDYGTSVNLGTYTEPYFTKDGVSILGGINFFENYDNSKSDTSSNYKRT 505
QY 377 AFSGGI---WVVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG 414
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 506 TYGSNVTLGFPVNNNSYVGLGHTYKNISNFALEYNRNLYIOSMKFKGNGIKTNDPFD-- 563
QY 415 NSHATMLTASWKROLLNNVLPENGHYLDGKIGTTLGTFILSSTALIRTSARAGYFFTPEN 474
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 564 -----PFCWNNNSLNRGYFTKG--VKASLGRVTPICSDKNKYKLSADVOGYFLDR 615
QY 475 KKLGTFTIRQAGTYTARDNADVPVSLMFRSGGASSVRGYELDSIGLAGPN----- 525
Db 616 DHRWVWSAKASAGYANGFGNKRLLPFYQTYTAGGIGSLRGFAYGSI---GPNATYAEHNG 672
QY 526 -----GSLVPERALLVGSLEYOLP-----FTRT-----L 549
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 673 TFNKKISSDVIGGNATITASAEILVPTFPVSDKSONTVRTSLFVDASVWNTKWSKRNGL 732
QY 550 SGAVFHDMDGAAANFKRMKLNKHSGLGVRWFPSPFAPFSFDIA-----YHSDKKIRWHIS 604
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 733 ESKVLKLDPDYG---KSSRIRASTGVGFQWQSPIGPLVFSYAKPIKKYENDDVE-QQFQS 788
QY 605 LGTRF 609
Db 789 IGGSF 793
```

```
RESULT 26
AAR53757
ID AAR53757 standard; Protein; 792 AA.
XX
XX AC AAR53757;
XX
XX DT 08-DEC-1994 (first entry)
XX
XX DE H. Influenzae SB33 D15 sequence.
XX
XX KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
XX OMP; outer membrane protein; NTHi.
XX
XX OS Haemophilus influenzae SB33.
XX
XX PN W09412641-A.
XX
XX PD 09-JUN-1994.
XX
XX PF 23-NOV-1993; 93WO-CA00501.
XX
XX PR 23-NOV-1992; 92GB-0024584.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX YI Yang Y;
XX
XX WPI: 1994-200269/24.
DR N-PSDB; AAQ66201.
XX
XX Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
XX PS Disclosure; Fig. 1D; 161pp; English.
XX
XX CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
XX SQ Sequence 792 AA;
Query Match 6.8%; Score 210.5; DB 15; Length 792;
Best Local Similarity 21.5%; Pred. No. 3.1e-08;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps 21;
QY 134 IANVG--VAIIGDILSDGNLAERYRNALENKQPVGSDFDQDSWENSKTSVLGAVTRKGY 191
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 270 IGNLGGMSAELEPLLSALHLNDTFRS-----DIADVENAIKAKLG---ERGY 314
QY 192 PLAKLGNTRAAVNDP----TATVDLNVVVDGRPIAFGDFEITGTQRYPEQIVSGLARFQ 247
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 315 ----GNTTVNSVPDFDDANKTLAITFVDAGRRLTVQLRFEGNTVSADSTLRQEMRQ 369
QY 248 PGTPYDLDLDDFOQALEQNGHYSAGVQADFRLQG--DRVPKVSVTEVKRHKLETGI 305
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 370 EGTWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVVYKVKERTGTSINFGI 427
QY 306 RLDSYGLGKGIADYYNLFNKNKGYIGSVWDMKYETTTLAAGISQPR-----NTRGNVW 359
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 428 GYGTESGISVQASVKQDNFLGTGAANSVIACTKNDYGTSVNLGTYTEPYFTKDGVSILG 487
QY 360 TSNVSYNRS-TTONLEKRAFSGGI---WVVRDRAGIDARLG-----AEFLAE----- 402
: | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Db 488 FENYDNSKSDTSSNYKRTYGSNTLGLFPVNNENSYVGLGHTYNNKISNFALEYNNRNYI 547
 QY 403 -----GRKIPGSDIDLGNSHATMLTASWKRQLLNVLHPNGHYLDCKIGTTLGTFLSS 456
 Db 548 QSMFKNGIKTDFD-----FSGWYNLSNRGYFPTKG--VKASLGCRVTIPGSD 597
 QY 457 TALIRTSARAGYFTPTPENKKIKGTPIIRGQAGYTVARNADVPGLMFRSGGASSVRGYEL 516
 Db 598 NKYYKLSADVOGFPLDRDLHWVYSASAGYANGFGNKRLLPFYQYTVTAGGIGSLRGFAY 657
 QY 517 DSIGLAGPN-----GSLVPERALLVGSLEYQLP-----FTRT----- 548
 Db 658 GSI---GPNAIYQGNKFNKISSDVIGGNAIATASAEILVPTPFVSDKSQNTVRTSLFV 714
 QY 549 -----LSGAVFHDMDGAAANFRKMKLKHGSLGVRWFSPLAPSFDA- 591
 Db 715 DAASVWNTWKMSKDNKNGLESNVLKDLDPYG---KSRTRASTGCVGFQWQSPGSPVVFYSAK 771
 QY 592 ----YGHSDKKIRWHISLGRF 609
 Db 772 PIKKYENDDVE-QQFSGSGSF 792

RESULT 27
 AAR53755
 ID AAR53755 standard; Protein; 797 AA.
 XX
 AC AAR53755;
 XX
 DT 08-DEC-1994 (first entry)
 XX
 DE H. influenzae b Eagen D15 sequence.
 XX
 KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; Hib.
 XX
 OS Haemophilus influenzae type b Eagen strain.
 XX
 PN W09412641-A.
 PD 09-JUN-1994.
 XX
 PF 23-NOV-1993; 93WO-CA00501.
 XX
 PR 23-NOV-1992; 92GB-0024584.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
 PI Yang Y;
 XX
 WPI; 1994-200269/24.
 DR N-PSDB; AAQ66199.
 XX
 PT Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 PS Disclosure; Fig. 1B; 161pp; English.
 XX
 CC Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AAQ66198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 797 AA;
 Query Match 6.6%; Score 209.5; DB 15; Length 797;

Best Local Similarity 20.0%; Pred. No. 3.7e-08;
 Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
 QY 61 DSEIKDWVEEHLPLITQOOEEVLDEKEQTGFLAEEAPONVKTMLRSKGYFSSKVSJLT----- 116
 Db 184 ESVSSSTLQEQMELQPDSSWKLGNKFEQAQFKDLQSIIRDYIYLNNGYAKAOKITKDVL 243
 QY 117 --EKDGA-YTVHIITPGPRTK-----IANVG--VAILCDILSDGNLAERYNALENWQOP 165
 Db 244 NDEATKYNVTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTPRS----- 296
 QY 166 VGSDFDSDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVPD-----TATVDLVNVVDSGRP 221
 Db 297 -----DIADVENAIKAKLG-----ERGYGSATVNSV-----PDFDANKTLAITLVVDAGR 343
 QY 222 IAFGDFEITCQRYPEQIVISGLARFQCPYDLDLILLDFQALAEQNGHYSASVQADPDR 281
 Db 344 LTVQLRFEGTVTSADSTLRQEMRQOEGTWYNSOLVELGKIRLDRDTGFFE--TVENRIDP 401
 QY 282 LQG--DRVPKVSVTEVKRHKLETGIRLDSYGLGGKIAYDYNYLNFKNKGYIGSVWMDMK 339
 Db 402 INGSNDEVVYVKERNVTGSIINFGYGTSGISYQASVKQDNPLGCTGAANVSIAGTKND 461
 QY 340 YETTLAAGISQPR-----NYRGNVWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
 Db 462 YGTSVNILGYTBPYFTKDCVSLGGNVFPENYONSKSDTSSNYKRTYGSNVTLGFPVNENN 521
 QY 390 GIDARLG-----AEFLAE-----GRKIPCSIDIDLGNSHATMLTASWKROLL 430
 Db 522 SYVYGLGHTYNNKISNFALEYNNRNYIOSMKFKNGIKTDFD-----FSFGWYNLSL 573
 QY 431 NNVLHPENGHYLDCKIGTTLGTFLSSTALIRTSARAGYFTPTPENKKIKGTFTIIRGQAGYTV 490
 Db 574 NRGYPTPKG--VKASLGCRVTIPGSDNKKYKLSADVOGFPLDRDLHWVVSASAKASYAN 631
 QY 491 ARDNADVPGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531
 Db 632 GFGNKRLLPFYQYTVTAGGIGSLRGFAYSI---GPNAIYAEGNGSGTGTFFKKISSDVIGG 688
 QY 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFIRM----- 567
 Db 689 NAIATSAELIVPTPFVSDKSQNTVRTSLFVDAASVWNTWKMSKDNKNGLESVDLKRLLPDYG 748
 QY 568 ---KLKHGSGLGVRWFSPLAPSFEDIA-----YGHSDKKIRWHISLGRF 609
 Db 749 KSSRIRASTGVGFQWQSPGIGPLVFSYAKPIKKYENDDVE-QQFSGSGSF 797

RESULT 28
 AAR53756
 ID AAR53756 standard; Protein; 797 AA.
 XX
 AC AAR53756;
 XX
 DT 08-DEC-1994 (first entry)
 XX
 DE H. influenzae b Minn A D15 sequence.
 XX
 KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; NTHI.
 XX
 OS Haemophilus influenzae type b Minn A strain.
 XX
 PN W09412641-A.
 PD 09-JUN-1994.
 XX
 PF 23-NOV-1993; 93WO-CA00501.
 XX
 PR 23-NOV-1992; 92GB-0024584.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.

Qy 340 YETTLAAGISQPR-----NVRGNWTSNVSYNRS-TTONLEKRAFSGGI---WYVRDRA 389
 Db 462 YCTSVNLGYTEPYFTKDCVSLGNGVFFENYDNSKSDTSSNYKRTTYGNSVTLGFPVNENN 521
 Qy 390 GIDARLG-----AEFLAE-----GRKIPCSIDLGNSHATMLTASWKROLL 430
 Db 522 SYVGLGHTYWKISNFALAYNRNLYIOSMKFKGNGIKTNDPD-----FSFGWYNNSL 573
 Qy 431 NNVLHPENGHYLDKIGTTLGTFLSSTALIRTSARAGYFFPPFNKLGCTFIIRGOAGYTV 490
 Db 574 NRGYFPTKG--VKASLGRVTIPGSDNNKYKLSADVGFFPLDRDHLWVYSKASAGYAN 631
 Qy 491 ARDNADVPGLMFRSGGASSVRGVELDIOGIAGPN-----GSLVPE 531
 Db 632 GFGNKRLLPFYQTYTAGGIGSLRGFPYGSII--GPNAYIAEYNGSGTGCTFKKISSDVIGG 688
 Qy 532 RALLVGSLEYOLP-----FTRLSGAVFHD-----GDAAANFKRM----- 567
 Db 689 NAITASPELIVPTPFVSDKSQNTVRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
 Qy 568 ---KLKHGSGLGVRWFSPAPFSFDIA-----YGHSDKKIRWHISLGTFR 609
 Db 749 KSSRIRASTGVGFQFQSPIGLPLVSYAKPIKKYENDDOVE-QQFSGSGSF 797

RESULT 30
 AAW55729
 ID AAW55729 standard; Protein: 925 AA.
 XX
 AC AAW55729;
 XX
 DT 13-JUL-1998 (first entry)
 XX
 DE H. pylori ORF 05ep10815_4719175_c1_115 cell envelope OMP.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW cell envelope outer membrane protein; OMP.
 XX
 OS Helicobacter pylori.
 XX
 PN W09737044-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US05223.
 XX
 PR 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX
 DR WPI: 1997-503122/46.
 DR N-PSDB: AAV25138.
 XX
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 XX
 PS Claims 14,80; Page 999-1001; 1145pp; English.
 XX
 CC This sequence represents a Helicobacter pylori cell envelope outer
 CC membrane protein having a terminal Phe residue.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55799) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 XX
 SQ Sequence 925 AA;
 Query Match 5.5%; Score 174; DB 18; Length 925;
 Best Local Similarity 21.0%; Pred. No. 3.9e-05;
 Matches 123; Conservative 72; Mismatches 202; Indels 188; Gaps 23;
 Qy 186 VTRKGYPLAKGNTRAAVNPDTATVDLNVV-----DSGRPIAGDPFITGORYPEQI 239
 Db 368 IADKGYAF-----AVVKPDLDKDEKNGLVKVIYRIEVDGMVHINDVITISGNQRTSDRI 420
 Qy 240 VSGLARFQPGTPYDLDLLDFQALQNGHYSGASVOADFRLQGDVRVPKVSVTEVKRH 299
 Db 421 IRRELLGPKDKYKYNLTKLRSNSLRLLGFFS--KVIEKRVNSSLMDLVSYVEEGRTG 478
 Qy 300 KLETGIRLDSEYGLGKIAIDYDYNLFNKGYIGSVWMD-----MDKYETTLAAG- 347
 Db 479 QLQFGLGVSYGGLMLNGSVSRNLFGTQGSMSLYANIATGGRSYPCMPKAGRMFAGN 538
 Qy 348 --ISQPR-----NYR-----GNTWTSNVNSTRSTQ 371
 Db 539 LSLNPRIFDSWYSSTINLYADYRISYQYIQGGGFGVNVGRLGNRTHVSLGYNLNVTK 598
 Qy 372 NLEKRAFSGGIWYVDRAGID-----ARLGAEEFLAEGRK--IPGSDIDLG--- 414
 Db 599 LL---GFSSPL-YNRYYSVNEVASPRQCSTPASVINRLSGGRTPLVPESCSPGALT 654
 Qy 415 -----NSHATMLTASWKROL-----LNNVLHPENG-----HYLDCKIGTTLGTFLSST 457
 Db 655 SPEIKGIWDRDYHTPITSSFTLDVSYDNTDDYFPRNCVIFSSYATMSGLPSSGTLNSWN 714
 Qy 458 AL---IRTSARAGYFFTPENKKLGTFI-----IRGQAGYTVARDNAD--VPISGLMFRSGG 507
 Db 715 GLGNVRNTKVYGFKAAYHHQYLLDLIARFKTGGY-IFRYNTDDYLPNSTFYMG 773
 Qy 508 ASSVRGYELDSIGLAGPNGSVLPE-----NGSITPKDERGLWLGDDGIFTASTELSYGLKAAKMLAWFFD 824
 Db 774 VTVVRGFR-----NGSITPKDERGLWLGDDGIFTASTELSYGLKAAKMLAWFFD 824
 Qy 532 -----RALLVGSLEYQLPFTTLTSLGAVFHDMDAAANFKRMKLGHSGLGVRWFSPLAP 585
 Db 825 FGLTFTKTPTRGSFFYNAPTTT---ANFKDYGVVVGAFERATWRASTGLQIEWISPMGP 880
 Qy 586 --FSFDIAY-----GHSDDKKIRWHISLGTFR 609
 Db 881 LVLIFFPIAFFNQWDGNGKCKKGLCFNPNMNDYTQHFESFSGTFR 925

RESULT 31
 AAY17183
 ID AAY17183 standard; Protein: 925 AA.
 XX
 AC AAY17183;
 XX
 DT 03-AUG-1999 (first entry)
 XX
 DE H. pylori outer membrane polypeptide.
 XX

KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX cellular immune response.
OS Helicobacter pylori.
XX W09921959-A2.
XX 06-MAY-1999.
XX 28-OCT-1998; 98WO-US22883.
XX 17-DEC-1997; 97US-0993001.
XX 28-OCT-1997; 97US-0959131.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
XX WPI; 1999-326698/27.
XX N-PSDB; AAX75802.
XX Cellular vaccine against Helicobacter pylori
XX Claim 7: Page 243-247; 352pp; English.
XX The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
CC (OMPs) AAY17160 to AAY17218.
XX
SQ Sequence 925 AA;

Query Match 5.5%; Score 174; DB 20; Length 925;
Best Local Similarity 21.0%; Pred. No. 3.9e-05;
Matches 123; Conservative 72; Mismatches 202; Indels 188; Gaps 23;

QY 186 VTRKGYPLAKGNTRAAVNDPTATVDLVVV-----DSGRPIAFGDFEITGTQRYPEQI 239
Db IADKGYAF-----AVVPPDLKDEKGLVKVIYRIEVDGMVHINDVIISGNQRTSDRI 420
QY 240 VSLARFQPTPYDLDLLDFQALQNGHYSCASVQADFDRLOGDRVPVKVSYTEVKRH 299
Db IRRELLGPKDKYNLTKLRNSENLSRLRGFFS--KVKIEKRVNSSLMDLLSVVEGRGTG 478
QY 300 KLETGIRLDSEYGLGKIAVDYVNLFNKGYIGSVWVD-----MDKYETTLAAG- 347
Db QLOFGLGYSGYGLMLNGSVSENLFTGQSMSLYANIATGGGRSYPGMPKAGRPFAGN 538
QY 348 --TSQPR-----NRYR-----GNYTWSNVSYNRSTTQ 371
Db LSLTNPRIFDSWYSSSTINLYADYRISVQYIQQGGGCVNVGRMLGNRTHVSLCYNLNVTK 598
QY 372 NLEKRAFSGGIWVRDAGID-----ARLGAFLAEGRK--IPCSDDILG---- 414
Db LL---GFSSPL-YNRYSSVNEVASPRQCSTPASVINRLSGGRTPLVPESCSGPAIT 654
QY 415 -----NSHATMLTASMKROL-----LNNVLHPENG-----HYLDGKIGTTLGTLFSS 457
Db SPEIKGIWDRDYHTPTSSFTLDVSDNTDDYFPNGVIFSSYATMSGLPSSGTLNSWN 714
QY 458 AL---INTSARAGYFTTPENKKGITFTI-----IRGAGYTVARDNAD--VPSGIMFRSGG 507
Db GLGNVNRNTKVKYKFAAYHLLQYLLDILAREKTOGGY-IFRYNTDDYLPNSTFVMGG 773
QY 508 ASSVRGYELDSIGLAGNGSVLPE----- 531
Db VTTVRGFR-----NGSITPKDEFGLWGGDIGFTASTELSYGVKKAAMRLAWFFD 824

QY 532 -----RALLVGSLEYOLPFTRTLTSAGVFDHMGDAANFKRMKLGKSGLVGRWFSPLAP 585
Db 825 FGFLTEKTPTRGSSFFYNAPTTT-----ANFKDYGVVGAGFERATWRASTGLQIEWISPMGP 880
QY 586 --FSFDIAY-----GHSDKKIRWHISLGTGRF 609
Db 881 LVLIFFIAFFNMGWDGNGKCKGLCFNPNMNDYTOHFEFSMGTRF 925

RESULT 32
ABB90599
ID ABB90599 standard; Protein; 790 AA.
XX
AC ABB90599;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chlamydia pneumoniae cp6576 protein, SEQ ID NO:147.
XX
KW Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT Peptide 1..26 /label= Signal_peptide
FT Protein 27..790 /note= "Mature protein"
FT
XX
PN W0200202606-A2.
XX
PD 10-JAN-2002.
XX
PF 03-JUL-2001; 2001WO-1B01445.
XX
PR 03-JUL-2000; 2000GB-0016363.
PR 11-JUL-2000; 2000GB-0017047.
PR 21-JUL-2000; 2000GB-0017983.
PR 07-AUG-2000; 2000GB-0019368.
PR 18-AUG-2000; 2000GB-0020440.
PR 14-SEP-2000; 2000GB-0022583.
PR 10-NOV-2000; 2000GB-0027549.
PR 22-DEC-2000; 2000GB-0031706.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Ratti G, Grandi G;
XX
DR WPI; 2002-154726/20.
DR N-PSDB; ABL91257.
XX
PT Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -
XX
PS Claim 1; Page 113; 364pp; English.
XX
CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of

```
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX
SQ Sequence 790 AA;

Query Match 5.5%; Score 173.5; DB 23; Length 790;
Best Local Similarity 20.7%; Pred. No. 3.4e-05;
Matches 148; Conservative 92; Mismatches 266; Indels 209; Gaps 33;

Qy 22 PAADLSENKAAGFALFNKSPDTSVKLPKFPVRIDTQDSEIKDMVEEHLPLITQ--Q 79
Db 87 PKVEFSEK--TNIALHLIAKPSIRNIHSG-----NQVVPKHILKLTQIYR 132
Qy 80 EVDLDKQGTGLABEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVH-----ITPGPRTK 133
Db 133 NDLFEREK--FL--KGLDRLTYVYLRKGYFASVDYSLEHNOEKGHIDVLKINEGCGK 188
Qy 134 IANYGVAILG-----DILSDGNLA-----EYY 155
Db 189 IKQLTFSGISRKSEKSDIQEFTQKHSHTTTSWFTGAGLYHPDIVQEDSLAITNVLHNGY 248
Qy 156 RNALNNQOQVGSDFDQDSWEN-----SKTSVLGAVTRKGYPL----- 193
Db 249 ADAI-----VNSHYDLDDKGNILLYMDIDRGSRYTLGHVHIQGFVLPKRRLIEKQSQVG 302
Qy 194 -----AKLG--NTRAAV-----NPDATVTLNVDVLFIPHATRPIDYDVYVSEGSYPKVG 226
Db 303 PNDLYCPDKIWDGAHKIKQTYAKGYINTNVDVLFIPHATRPIDYDVYVSEGSYPKVG 362
Qy 227 FEITGTORYPEQIVSGLARFQPGTPYDLDLLDFQQALEQNGHYSGASV-----OA 277
Db 363 IKITGNTHTKSDVILHETSLPGDTPNRLKLEDEQRLNRTGYFQSVSVTVRSQOLDPMG 422
Qy 278 DFDRLQGRDVPVKVSVTE-----VKRHKLETGIRL--DSEYGLGKGIADYVNLNKG 328
Db 423 NADQYRDI FVEVEKTTTGNLGLFLGFSLSLNLFGGIELSESNFLDGA-----RNIFSKG 477
Qy 329 Y-----IGSVVWMDKYETTTLAAGISOPRNYRGNVYTSNV-----SYNRST 369
Db 478 FRCRLGGEGEHLFLKANFGDKVDTYTLKWT-----RP-HFLNTPWILGIELDKSINRAL 529
Qy 370 TONLEKRAFSGGI--WVYRD--RAGIDARLGAFLAEGRK--IPGSDIDLGNSHATMLTA 423
Db 530 SKDVAVOTYGGNVSTYILNHLKYGFLYRGSTSLHKKRKFLLGPNIDSKNGFVSAAG 589
Qy 424 SWKRLQLNVLHPENG-----HYLDGKIGTTLCTFLS--STALIRTSARAGYFFTPENKK 476
Db 590 NLNTVDSVSPRTPITGIRGGVTFVSGLGGYHFTKLSLSNYSIYRKLTR----- 639
Qy 477 LGTFIIRQAGYTVARDNA-----DVPGLMFRSGGASSVRGYELDSIG----LAGPNGSVL 529
Db 640 -GILKIRGEAQFIRPYSNTTAEQVPSERFPLGGETTVRGYKSFIIIGPKYSATEPQGL- 697
Qy 530 PERALLVGSLEYQLPFR--TLSGAVFHDMDGAAANFKRMKLK---HSGSLGVWR 579
Db 698 --SSILLI--SEEFQVPLIRQPNISAFVLDGFGVGLQEVKISKDLRSSAGFGLRF 749

RESULT 33
AA34896
ID AA34896 standard; Protein; 795 AA.
XX
AC AA34896;
XX
AC AA34896;
XX
DT 13-SEP-1999 (first entry)
XX
XX Chlamydia pneumoniae transmembrane protein sequence.
DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
```

```
XX Chlamydia pneumoniae.
OS
XX WO9927105-A2.
PN
XX 03-JUN-1999.
PD
XX 20-NOV-1998; 98WO-IB01890.
PF
XX 04-NOV-1998; 98US-0107078.
PR
XX 21-NOV-1997; 97FR-0014673.
XX (GEST ) GENSET.
PA
XX Griffiths R;
PI
XX WPI; 1999-357842/30.
XX
XX Genome sequence of Chlamydia pneumoniae
PT
XX Page 832-834; Disclosure; 1912pp; English.
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY34584) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 795 AA;

Query Match 5.5%; Score 173.5; DB 20; Length 795;
Best Local Similarity 20.7%; Pred. No. 3.4e-05;
Matches 148; Conservative 92; Mismatches 266; Indels 209; Gaps 33;

Qy 22 PAADLSENKAAGFALFNKSPDTSVKLPKFPVRIDTQDSEIKDMVEEHLPLITQ--Q 79
Db 92 PKVEFSEK--TNIALHLIAKPSIRNIHSG-----NQVVPKHILKLTQIYR 137
Qy 80 EVDLDKQGTGLABEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVH-----ITPGPRTK 133
Db 138 NDLFEREK--FL--KGLDRLTYVYLRKGYFASVDYSLEHNOEKGHIDVLKINEGCGK 193
Qy 134 IANYGVAILG-----DILSDGNLA-----EYY 155
Db 194 IKQLTFSGISRKSEKSDIQEFTQKHSHTTTSWFTGAGLYHPDIVQEDSLAITNVLHNGY 253
Qy 156 RNALNNQOQVGSDFDQDSWEN-----SKTSVLGAVTRKGYPL----- 193
Db 254 ADAI-----VNSHYDLDDKGNILLYMDIDRGSRYTLGHVHIQGFVLPKRRLIEKQSQVG 307
Qy 194 -----AKLG--NTRAAV-----NPDATVTLNVDVLFIPHATRPIDYDVYVSEGSYPKVG 367
Db 308 PNDLYCPDKIWDGAHKIKQTYAKGYINTNVDVLFIPHATRPIDYDVYVSEGSYPKVG 367
Qy 227 FEITGTORYPEQIVSGLARFQPGTPYDLDLLDFQQALEQNGHYSGASV-----OA 277
Db 368 IKITGNTHTKSDVILHETSLPGDTPNRLKLEDEQRLNRTGYFQSVSVTVRSQOLDPMG 427
Qy 278 DFDRLQGRDVPVKVSVTE-----VKRHKLETGIRL--DSEYGLGKGIADYVNLNKG 328
Db 428 NADQYRDI FVEVEKTTTGNLGLFLGFSLSLNLFGGIELSESNFLDGA-----RNIFSKG 482
Qy 329 Y-----IGSVVWMDKYETTTLAAGISOPRNYRGNVYTSNV-----SYNRST 369
Db 483 FRCRLGGEGEHLFLKANFGDKVDTYTLKWT-----RP-HFLNTPWILGIELDKSINRAL 534
Qy 370 TONLEKRAFSGGI--WVYRD--RAGIDARLGAFLAEGRK--IPGSDIDLGNSHATMLTA 423
```

Db 535 SKDYAVQYGGNVTYILNEHLKYGLFYRGSTSLHEKRKFLGPNIDSNKGFVSAAGV 594
QY 424 SWKRQLNNVLPENG-----HYLDGKIGTTLGTFLS-STALIRTSARAGYFFTPENKK 476
Db 595 NLNYDSVDSPTTTGIRGGVTEVSGLGTYIFTKLSLSSSYRKLTRK----- 644
QY 477 LGTFIIRGOAGYTVARDNA---DVPGLMFRSGGASSVRGYELDSIC---LAGPNSVYL 529
Db 645 -GILKIKGEAQFIKPYSTNTTAEGVPVPSERFFLGGETTVRGYKSPFIIGPKYSATEPOGGL- 702
QY 530 PERALLVGSLEYQLPSTR--TLSCAVPHDMDGDAANFKRKKLK---HGSLGLGVRR 579
Db 703 --SLLLI-SEFQYPLIRQNIASFVLDGFGYGLQEKYKISLRLSRAGFLRF 754
RESULT 34
AAW71477
ID AAW71477 standard; Protein; 847 AA.
AC AAW71477;
XX
DT 09-NOV-1998 (first entry)
XX
DE Helicobacter polypeptide GHPO 107.
XX
KW GHPO 107; infection; therapy; diagnosis; vaccine; gastritis;
KW ulcer.
XX
OS Helicobacter pylori.
XX
PN W09821225-AI.
XX
PD 22-MAY-1998.
XX
PF 14-NOV-1997; 97NO-US21353.
XX
PR 29-JUL-1997; 97US-0902615.
PR 14-NOV-1996; 96US-0749051.
PR 01-APR-1997; 97US-0831309.
PR 01-APR-1997; 97US-0833457.
PR 01-APR-1997; 97US-0834705.
PR 24-JUN-1997; 97US-0881227.
XX
(HUMA-) HUMAN GENOME SCI INC.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX
WPI: 1998-297855/26.
DR N-PSDB; AAV52012.
DR
PT Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
XX
PS Claim 1; Page 86-88; 362pp; English.
XX
CC This claimed Helicobacter pylori polypeptide, designated GHPO 107,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC AAW71474-W71558) are claimed, as well as isolated polynucleotides
CC (see AAV52009-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transformed or transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention and which can be
CC used to prevent or treat Helicobacter infection; therapeutic and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or monospecific antibodies; methods for
CC detecting the presence of Helicobacter in samples using e.g.
CC the polypeptides or monospecific antibodies; and methods for

CC purifying the polypeptides by antibody-based affinity
XX chromatography.
SQ Sequence 847 AA;
Query Match 5.4%; Score 172.5; DB 19; Length 847;
Best Local Similarity 20.0%; Pred. No. 4.5e-05;
Matches 163; Conservative 91; Mismatches 230; Indels 333; Gaps 37;
QY 76 TQOQEEVLDEKQGF-----LAEAPDNVKTMLRSKYFSSKVS-L-TEK--DCAYT 123
Db 81 TEKEKQGL-KSOMGIKKGDTFDEQKLEHAKTALKTALEGGYGVVVEVTEKVESEGALL 139
QY 124 VHTPGPRTKIANVQVAI-LGDILSDGNLAEEYRNALENWQPVQSPDFDQDSW----- 175
Db 140 IVFD-----VNRGDSIYIKQSIYEGS-AKLKRMIESLSANKORDFMGMWGLNDGKL 191
QY 176 -----ENSKTSVLGAVTRKY----- 191
Db 192 RLDQLEYDSMRIQDVVMRGYLDHAHISPPFLKTDFTSHDAKLHYKVKEGIQYRISDILIE 251
QY 192 -----PLAKL-----GNTRAAVNPDTATVDLVN 214
Db 252 IDNPVVPKLTLEKALKVKRKDVFNIEHLRADAQILKTEIADKGYAFVAVKPDLDKDKNG 311
QY 215 VW-----DSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPVDLDDLDFQOALBQNG 268
Db 312 LVKVIVRIEVDMAVYINDVILISGNQRTSDRIIRRELLGPKDKYNLTKLNSNSLRIG 371
QY 269 HYGASVQADFDRIQGDVRPVKSVTEVKRHKLETGIRLSEYGLGCKIAYDYNNLFNKG 328
Db 372 FFS--KVKIEKRVNSLMDLLVSEBGRGTQLQFGLGYGSGGLMLNGSVSRNLPGTG 429
QY 329 YIGSVVND-----MDKYETTLAAG---ISQPR-----NYSR----- 355
Db 430 QSMSLYANIATGGRSYPGPKGAGRMFAGNLSLTNPRIFDSWYSSTINLYADYRISYQY 489
QY 356 -----GNWTSNVSYNRSSTONLEKRAFSGGIWVVRDRAGID----- 392
Db 490 IQOQGGFGVWVGRMLGNRTHVSLGYNLVNFKLL---GFSSPL-YNRYSSVNEVVSPPQC 545
QY 393 ---ARLGAEEFLAEGR-----KIPGS-----DIDLGN- 415
Db 546 STPASVIINRLSGKTKPLQPESSCSGCAITTSPEIRGIWDRDHYTPITTSFTLDVSYDNT 605
QY 416 -----SHATM--LTASWKRQLLNNVLHPENGHYLDGKIGTT--LGTF----- 453
Db 606 DDYFFPRNGVIFSSYATMSGLPSS-----GTLNSWNG--LGNVNRNTKYGKFAAYHH 656
QY 454 LSSTALIRTSARAGYFFTPENKKLGTFTIRGOAGYTVARDNAD--VPSGLMFRSGGASSV 511
Db 657 LQKYLILDLIAR-----FKTQGGY-IFRYNTDDYLPINSTFFYMGVTTV 699
QY 512 RGYELDSIGLAGPNSVLP--ERALLVGS-----L 539
Db 700 RGRF-----NGSVTPKDFGLWLGDDGIFTASTELSYGLVLAAMKRLAWPFDFGFL 750
QY 540 EYQLP-----FRTLSCAVPHDMDGDAANFKRKKLHGSGLGSLGVRFWSPLAP--FSFDIA 591
Db 751 TFKTPTRGSPFYNAVPTTANFKDYGVICAGFERATWRASTGLQIEWISPMGLVLIPPIA 810
QY 592 YGH-----SPKKIR-----WHISLGTFR 609
Db 811 FFQWGDGNGKCKKGLCFNPNMDDYTQHFESMGTRF 847
RESULT 35
AAB46311
ID AAB46311 standard; Protein; 916 AA.
AC AAB46311;
XX
DT 05-APR-2001 (first entry)

XX H. pylori HPS120 protein.
 DE Microbial infection; antibacterial; Helicobacter pylori infection;
 KW vaccine; screening.
 XX Helicobacter pylori.
 OS WO200073502-A2.
 XX 07-DEC-2000.
 PD
 XX 31-MAY-2000; 2000WO-EP05024.
 PF
 XX 31-MAY-1999; 99DE-1024965.
 PR 17-JUN-1999; 99DE-1027740.
 PR 21-JUL-1999; 99DE-1034029.
 XX (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 XX
 PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;
 XX WPI; 2001-049948/06.
 DR N-PSDB; AAF25588.
 XX
 XX Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 PT essential genes in defective mutants -
 XX
 XX Claim 37; Page 236-239; 366pp; German.
 PS
 XX This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) (which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.
 XX
 XX Sequence 916 AA;
 SQ
 Query Match 5.4%; Score 172.5; DB 22; Length 916;
 Best Local Similarity 20.0%; Pred. No. 5.1e-05;
 Matches 163; Conservative 91; Mismatches 230; Indels 333; Gaps 37;
 QY 76 TQOQEEVLDRKQGTG-----LAEEAPDNVTKLRSGYFSSKYSVL-TEK--DGYAT 123
 DB 150 TEKERDGL-KSQMGIKKGDTFDEOKLEHAKTALKTALEGQGYGVSVVEVTRKEVSEGALL 208
 QY 124 VHTPGPRTKIANVGVAI-LGDIILSDGNLAERYNNALENMQQVPVGSDFDQSW----- 175
 DB 209 IVFD-----VNRGDSIYIKQSIYEGS-AKLKRMIETSLANKORDFMGMWGLNDGKL 260
 QY 176 -----ENSKTSVLGAVTRKY----- 191
 DB 261 RLDOLEYDSMRIQDYVMRRGYLDIAHISPFLLKTFSTHDAKLHYKVKREGIQYRISDILIE 320
 QY 192 -----PLAKL-----GNTRAVNPDTATVDLNV 214
 DB 321 IDNPVPLKLEKALKVKRKDVFNIEHLRADAQILKTEIAOKGYAFVAVKRPDLDRKDEKNG 380

QY 215 VV-----DSGRPIAFGDFEITGTQRYPEQIVSGLARFPQCTPYDLDLLDFQOALEQNG 268
 DB 381 LVKYIYREVDVMYVINDVIISGNQRTSDRIIRRELLGPKDKYNLTKLRSNLSRLRG 440
 QY 269 HYGASVQADFDRLQGRVPVKVSVTEVKRHKLETGRDSEYGLGGKIAIYDYNLFPNG 328
 DB 441 PFS--KVKIEKRVNSSLMDLLSVSEEGRTQLOGLCYGSGYGLMLNGSVSERNLFCTG 498
 QY 329 YIGSVVWD-----MDKYETTLAAG---ISQPR-----NVR----- 355
 DB 499 QMSLYANIATGGRSYPGMPKAGRMFAGNLSLTNPRIFDSWYSSTINLIADYRISYQY 558
 QY 356 -----GNYVTSNVSYNRSTTQNLKRAFGSGGIWYVRDRAGID----- 392
 DB 559 IQGGGFCVNVGRMLGNRTHVSLGYNLNVTKLL---GFSSPL-YNRYYSVNEVSVPRQC 614
 QY 393 ---ARLGAELAEGR-----KIPGS-----DIDLGN- 415
 DB 615 STPASVIINRLSGGKTPLOPESCSPGAITTSPEIRGIWRDHYHTPTTSSTFLDVSVDNT 674
 QY 416 -----SHATM--LTASWKROLLNNVLHPENCHYLDGKIGTIT--LCTF----- 453
 DB 675 DDYFPRNGVIFSSYATMSGLPSS-----GTLNSMWG---LGGNVRNTKYVGKPAAYVHH 725
 QY 454 LSSTALRTSARAGYFFTPENKKLGTFTIRGOAGYTVARDNAD--VPSGLMFRSGGASSV 511
 DB 726 LQYLLIDLIAR-----FKTQGGY-IPRYWTDYDYLPLNSTFYMGVTV 768
 QY 512 RGYELDSTGLAGPNSVLP--ERALLVGS-----L 539
 DB 769 RGFR-----NGSVTPKDFGLWLGCGDIFTASTELSYGLKAAKMRWLAWPFDGFL 819
 QY 540 EYQLP-----FRTLSCAVPHDMGDAANFKRMKLHGSGLGUVWFESPLAP--FSPDIA 591
 DB 820 TFKTPTRGSPFYNPATVYNTANPKDYGVIGAGFERATWRASTGLQIEWISPMGLVLIPDIA 879
 QY 592 YGH-----SDKKIR-----WHISLGTGF 609
 DB 880 FFNQWGDGNGKKKGLCFNPNMDDYQTHFEFSMGTRF 916
 RESULT 36
 AAY69362
 ID AAY69362 standard; Protein: 787 AA.
 XX
 AC AAY69362;
 XX 19-JUN-2000 (first entry)
 DT
 XX Amino acid sequence of the CPN100111 polypeptide.
 DE
 XX CPN100111; Chlamydia infection; immune response; vaccine.
 KW
 XX Chlamydia pneumoniae.
 OS
 XX Key Location/Qualifiers
 FT Peptide 1..85
 FT /note= "signal peptide"
 XX
 XX WO200011183-A2.
 XX
 XX 02-MAR-2000.
 PD
 XX 18-AUG-1999; 99WO-1801449.
 PF
 XX 20-AUG-1998; 98US-0097187.
 PR 20-AUG-1998; 98US-0097188.
 PR 20-AUG-1998; 98US-0097189.
 PR 20-AUG-1998; 98US-0097190.
 PR 20-AUG-1998; 98US-0097195.
 PR 20-AUG-1998; 98US-0097196.
 PR 20-AUG-1998; 98US-0097197.

Qy 151 LAEYRNALENWQOPV--GSDFDQD-----SWENSKTSVLGAVTRKGYPLAKLGNTRAAYN 204
 Db 200 LIDVLA-IKDNKINPLSKADRYTQEKLVTSLENLRKYLNA-----GFVRFEIKDAKLIN 254
 Qy 205 PDTATVDLVNVVDSGRPIAFGDFEITGQRYRPEQIVSGLARFOPCTPYDLDLDDLFQOAL 264
 Db 255 EDKNRIFVEISLHGEQYRFQOTQFLGNLTYYTQAELEALLKFAEEGFS-----QAML 307
 Qy 265 EQ-----NGHY-----SGASVQADFDRLOGDRVPVKV-----SV 293
 Db 308 EOTTNNISTKFGDDGYVYQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQ 367
 Qy 294 TEVVRHKLETGIRLSEVGLGCKIAYDYNNLFNKG-----IGSW 334
 Db 368 DEVLRRMR---OLEGALASNQIKLSRARLMRTGFFKHVTVDTRPVNSPDQVDVNFV 424
 Qy 335 WDMKYEETTLAAGISQ-----PRNYRGNYWTSNYSNRSTTONLEKRAFSGGIW 383
 Db 425 EEQPSGSGSTIAAGYSQSGGVTFQDVSONNFMGTGCKHVNASFSETRREVYSLGMTNPFY 484
 Qy 384 -----YVRDRAG-----ID--ARLGAELFLAEGKIPG 408
 Db 485 TVNGVSQSLSGYRKTYDKNKNISNYVLDYSGSLSYGYPIDENQISFGLNADNTKLHG 544
 Qy 409 SDIDLGNSHATMLTA-----SWKROLLNNVLPHPENG--H 440
 Db 545 GRF-MGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVFTQGM 603
 Qy 441 YLDGKIGTTLGTLFSLSTALIRTSARAGYFFTPENKLGTFIIRGOA--GYTVARDNADVP 498
 Db 604 SVDLTVG-----FGDKTH--QKVYVQGNIRPFYIKK---SVLRGYAKLGY-----GNNLP 648
 Qy 499 SGLMFRSGASSVRGYELDSIGLAGPN-----GSLVPERALLVGSLEYQL 543
 Db 649 FYENFYAGGYGVRGYDQSSL--GPRSQAYLTARRGOOTTLGEVVGNGALATFGSELIL 705
 Qy 544 PF-----TRTL-----SCAVFHDWG-----DAAANFKR----- 566
 Db 706 PLPFKGDWIDQVRPVIFIEGGQVFDTTGMDKQTDLTQFKDPQATAEONAKAANRPLLTQ 765
 Qy 567 -MKLKHGSLGVRWFSPLAPESFDIA 591
 Db 766 DKQLYSAGVATWTPICPLISYA 791

RESULT 38

AAY44391

ID AAY44391 standard; Protein; 813 AA.

AC AAY44391;

XX 14-MAR-2000 (first entry)

XX M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.

DE BASB027; OM95; outer membrane protein; otitis media; treatment;
 KW diagnosis; bacterial infection.
 XX Moraxella catarrhalis.

XX WO9963093-A2.

XX 09-DEC-1999.

XX 31-MAY-1999; 99WO-EP03822.

XX 03-JUN-1998; 98GB-0011945.

PR 08-MAR-1999; 99GB-0005304.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Vinals-Bassols C;

PI

XX

DR WPI: 2000-105700/09.
 DR N-PSDB; AAZ29551.

XX Novel BASB027 polynucleotide and polypeptides from Moraxella
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis
 PT media

XX Claim 1; Page 102-104; 109pp; English.

XX The present sequence is BASB027 polypeptide, which shows significant
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella
 CC catarrhalis strain Mc931 (ATCC 43617). BASB027 polynucleotide and
 CC polypeptide can be used for diagnosis and staging of disease, determining
 CC susceptibility to a disease and to prepare medicaments for treating M.
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
 CC used as probe for screening of genetic mutations, serotype, taxonomic
 CC classification or identification. BASB027 agonists, antagonists and
 CC antibodies may be used to prevent and/or treat bacterial infections.

XX Sequence 813 AA;

Query Match 5.2%; Score 165.5; DB 21; Length 813;
 Best local Similarity 19.4%; Pred. No. 0.00016;
 Matches 145; Conservative 96; Mismatches 260; Indels 245; Gaps 36;

Qy 55 VRIDTQSEIKDMVEEHLPLITQ--QOEEVLDRKE--QTGF-----LAEAPDNV 99
 Db 82 VOVYHQGRITTYQYTER-PLIAEINPEGNRLIPREGLOEGLKNAGLAVGQPLKQATVQMI 140
 Qy 100 KTMRL-----SKGYSSKVSLS--TEKDG---AYTVHITPGPRTKIANVGVAILGDLSDGN 150
 Db 141 ETELTNQVISQGYNTETVTKOTMLDGNRVKLDMTFAEGKPARVVDINI-IGNHFSDAD 199
 Qy 151 LAEYRNALENWQOPV--GSDFDQD-----SWENSKTSVLGAVTRKGYPLAKLGNTRAAYN 204
 Db 200 LIDVLA-IKDNKINPLSKADRYTQEKLVTSLENLRKYLNA-----GFVRFEIKDAKLIN 254
 Qy 205 PDTATVDLVNVVDSGRPIAFGDFEITGQRYRPEQIVSGLARFOPCTPYDLDLDDLFQOAL 264
 Db 255 EDKNRIFVEISLHGEQYRFQOTQFLGNLTYYTQAELEALLKFAEEGFS-----QAML 307
 Qy 265 EQ-----NGHY-----SGASVQADFDRLOGDRVPVKV-----SV 293
 Db 308 EOTTNNISTKFGDDGYVYQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQ 367
 Qy 294 TEVVRHKLETGIRLSEVGLGCKIAYDYNNLFNKG-----IGSVV 334
 Db 368 DEVLRRMR---OLEGALASNQIKLSRARLMRTGFFKHVTVDTRPVNSPDQVDVNFV 424
 Qy 335 WDMKYEETTLAAGISQ-----PRNYRGNYWTSNYSNRSTTONLEKRAFSGGIW 383
 Db 425 EEQPSGSGSTIAAGYSQSGGVTFQDVSONNFMGTGCKHVNASFSETRREVYSLGMTNPFY 484
 Qy 384 -----YVRDRAG-----ID--ARLGAELFLAEGKIPG 408
 Db 485 TVNGVSQSLSGYRKTYDKNKNISNYVLDYSGSLSYGYPIDENQISFGLNADNTKLHG 544
 Qy 409 SDIDLGNSHATMLTA-----SWKROLLNNVLPHPENG--H 440
 Db 545 GRF-MGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVFTQGM 603
 Qy 441 YLDGKIGTTLGTLFSLSTALIRTSARAGYFFTPENKLGTFIIRGOA--GYTVARDNADVP 498
 Db 604 SVDLTVG-----FGDKTH--QKVYVQGNIRPFYIKK---SVLRGYAKLGY-----GNNLP 648
 Qy 499 SGLMFRSGASSVRGYELDSIGLAGPN-----GSLVPERALLVGSLEYQL 543
 Db 649 FYENFYAGGYGVRGYDQSSL--GPRSQAYLTARRGOOTTLGEVVGNGALATFGSELIL 705
 Qy 544 PF-----TRTL-----SCAVFHDWG-----DAAANFKR----- 566
 Db 706 PLPFKGDWIDQVRPVIFIEGGQVFDTTGMDKQTDLTQFKDPQATAEONAKAANRPLLTQ 765

Qy	232	TQRTPEQIVSGLARFGQGT	PYDLDLLDFQOALEQNGHYSGAS---	VQADFRLQGD---	285	
		::	::	::		
Db	371	NTHTKHDVILHETSLFFG	QDTFYRLKLEDTETRLRNTGYFKSVSVYTVRSQ	LDPLDSNDLY	430	
		::	::	::		
Qy	286	-RVPKVSVTVEVKRHKLE	TGT-RLDSYGLGGKIAYDYNYLF-	-----NKGY-----	329	
		::	::	::		
Db	431	RDFVIEVKETGMLGLFL	GFSSIDHULFG-GAEIAESNFOLFQARNFLKKG	FKFSLRGGGE	489	
		::	::	::		
Qy	330	-----IG-----	SVVMDMDKY-ETTLAAGISQPRNYR----	GNVWTSNVSNRSTTQ	371	
		::	::	::		
Db	490	YLFLEKANLGDKVTDY	TVKWKTKPHELNTPIWILGVFLDKSINKALSKDY	SVDTYGGNISTTY	549	
		::	::	::		
Qy	372	NLEKRAFSGGIWW-----	VRDRAGIDARLGAELAEGRKIPGSDIDILGNSHAT	MLTA	423	
		::	::	::		
Db	550	ILNDK-LKYGMYYRG	SQTSLSLRKKTSSSNRLG-----	PDLSNKGFSVAAGLNV	598	
		::	::	::		
Qy	424	SWK-----	ROLLNNVLHPENGHYLDGKGIGTTIGT---	FLSSTALIFTSARAGY	468	
		::	::	::		
Db	599	LYDSIDNPRKPTMGIR	FLNKLVELSGLGTYQFTKL-TASGSIVRLUTPKV	LKKRAEAXF	657	
		::	::	::		
Qy	469	FPTPENKKGITFIIR	GOAGYTVARDNADVP	SGLMFRSGGASSVRGYELDSIG----	LAGP	524
		::	::	::		
Db	658	I-----KPGGTTA	QG-----	IPVSEFFLGGETTVRGYKPF	ILIGKFSPTPEP	700
		::	::	::		
Qy	525	NGSVLPERALLVGSLEY	OLPRTTSLGSAVHRDMGDAAANFKRMKLK---	HGSLGLVVR---	578	
		::	::	::		
Db	701	OGGL--SSLLTEEFQ	YPLISOPCINAFVFLDSGFIGIEEHTIRLKD	ICSSAGFLRFD	758	
		::	::	::		
Qy	579	-----WFSPLAP	585			
		::	::	::		
Db	759	MNVVPMILGWGM	PPRP	774		
		::	::	::		
		RESULT	40			
		ABG25558				
ID		ABG25558	standard; Protein; 371 AA.			
XX		AC	ABG25558;			
XX		DT	18-FEB-2002 (first entry)			
XX		DE	Novel human diagnostic protein #25549.			
XX		KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX		KW	food supplement; medical imaging; diagnostic; genetic disorder.			
OS		OS	Homo sapiens.			
XX		PN	WO200175067-A2.			
PD		PD	11-OCT-2001.			
XX		PF	30-MAR-2001; 2001WO-US08631.			
XX		PR	31-MAR-2000; 2000US-0540217.			
XX		PR	23-AUG-2000; 2000US-0649167.			
PA		PA	(HYSE-) HYSEQ INC.			
PI		PI	Drmanac RT, Liu C, Tang YT;			
XX		DR	WPI; 2001-639362/73.			
XX		DR	N-PSDB; AAS89745.			
XX		PT	New isolated polynucleotide and encoded polypeptides, useful in			
XX		PT	diagnostics, forensics, gene mapping, identification of mutations			
XX		PT	responsible for genetic disorders or other traits and to assess			
XX		PT	biodiversity.			
PS		PS	Claim 20; SEQ ID NO 55917; 103pp; English.			
XX		CC	The invention relates to isolated polynucleotide (I) and			
XX		CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,			

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, for forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 371 AA;

Query Match 5.1%; Score 160.5; DB 22; Length 371;
Best Local Similarity 23.9%; Pred. No. 0.00013;
Matches 77; Conservative 27; Mismatches 103; Indels 115; Gaps 13;
QY 355 RGNW-----TSNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKI 406
DB 16 RGYWDPNSDLCPTLDFSYKMWPLKN-----PLEQYYLVQG-GF 53
QY 407 PGSDIDLGNSHATMLTAS-----WKQLLNVLHPENGHYLDGKIGTTLGTFLLSSTA 458
DB 54 KRTDLNDESDSTLVASRYWDLSSGQRAI---NLWSLDHFQTQGEITNTMLFYPGVM 110
QY 459 LIRTSARAGYF-----FTPENKKLGT-----FIIRGQA 486
DB 111 ISRTSRGGLMPTWGDQSQRYSIDYSNTANGSDVDFSVFQAQNVWVIRTYDRHREFTVTRGT 170
QY 487 GYTVARDNADVPSCLMFERSGASSVRGYELDSICLAGPNSVLPERALLYCSLEYQLPET 546
DB 171 GWIETGDFDKVPPDLRFFAGGDRSIRGYKSIQPKLPNGDLKGASKLITGSLEYQYNVT 230
QY 547 -----RTLSGAVF-----HDMGDAANEX---RMKIK-HGSG 574
DB 231 GKCTKWQGEWLAPGEVRLAPGATLATPELVASCSTGLNGLAANFHAELRRLPWHGGA 290
QY 575 LGVR-----WFSPLAPFSFDI 590
DB 291 MKPRPVHLNTW-----EGFYFDL 308

Search completed: November 9, 2002, 01:18:45
Job time : 44 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 9, 2002, 01:24:34 ; Search time 57 Seconds

(without alignments)
3276.599 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLIGTRF 609

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09857669/runat_05112002_110008_7750/app_query.fasta_1.775
-DB-Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669 @Cn1_1_19 @runat_05112002_110008_7750 *NCPU=6 *ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213.5	6.7	2989	3	US-08-433-522A-9
2	213.5	6.7	2989	3	US-09-135-166-9
3	213.5	6.7	2989	4	US-08-942-046-9
4	210.5	6.6	2974	3	US-08-433-522A-7
5	210.5	6.6	2974	3	US-09-135-166-7
6	210.5	6.6	2974	4	US-08-942-046-7
7	209.5	6.6	2949	3	US-08-433-522A-1
8	209.5	6.6	2949	3	US-09-135-166-1
9	209.5	6.6	2949	4	US-08-942-046-1
10	209.5	6.6	2950	3	US-08-433-522A-5
11	209.5	6.6	2950	3	US-09-135-166-5
12	209.5	6.6	2950	4	US-08-942-046-5

13	209.5	6.6	2984	3	US-08-433-522A-3
14	209.5	6.6	2984	3	US-09-135-166-3
15	209.5	6.6	2984	4	US-08-942-046-3
16	209.5	6.6	2987	3	US-08-433-522A-55
17	209.5	6.6	2987	3	US-09-135-166-55
18	209.5	6.6	2987	4	US-08-942-046-55
19	134.5	4.2	1697	4	US-09-346-408-5
20	121.5	3.8	2280	4	US-09-346-408-3
21	117	3.7	1890	4	US-09-346-408-7
22	114.5	3.6	1854	6	5233391-4
23	114	3.6	9171	1	US-08-038-682-5
24	114	3.6	9171	1	US-08-302-832-5
25	114	3.6	9171	2	US-08-530-198-5
26	114	3.6	9171	2	US-08-469-880-5
27	114	3.6	9171	2	US-08-728-470-5
28	114	3.6	9171	2	US-08-617-697-5
29	114	3.6	9171	4	US-08-719-641-5
30	112.5	3.5	1524	1	US-08-021-601-9
31	112.5	3.5	1524	1	US-08-082-849B-9
32	112.5	3.5	1524	5	PCT-US94-01624-9
33	112	3.5	3286	2	US-08-363-124A-1
34	111.5	3.5	1791	1	US-08-565-386-10
35	111.5	3.5	19227	3	US-09-090-793-13
36	111.5	3.5	40138	3	US-09-090-793-12
37	111	3.5	5100	1	US-08-164-292B-15
38	111	3.5	5100	1	US-08-164-292B-17
39	111	3.5	5100	1	US-08-164-292B-19
40	111	3.5	5100	1	US-08-164-292B-21
41	111	3.5	5100	1	US-08-164-292B-23
42	111	3.5	5100	1	US-08-164-292B-25
43	111	3.5	5100	3	US-08-845-623-15
44	111	3.5	5100	3	US-08-845-623-17
45	111	3.5	5100	3	US-08-845-623-19

ALIGNMENTS

RESULT 1
US-08-433-522A-9
; Sequence 9, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2989 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 390..2768
 ; US-08-433-522A-9

Alignment Scores:

Pred. No.: 6,28e-13 Length: 2989
 Score: 213.50 Matches: 127
 Percent Similarity: 35.04% Conservative: 85
 Best Local Similarity: 20.99% Mismatches: 254
 Query Match: 6.73% Indels: 139
 DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-08-433-522A-9 (1-2989)

Qy 105 SerLysGlyTyrPheSerSerLysValSerLeuThr-----Glulys 118
 Db 1071 AATAATGGCTATGCCAAGCACAATACTAAACCGGATGTCAGCTAAATGATGAAAA 1130
 Qy 119 AspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys----- 133
 Db 1131 ACAAAAGTTAATGTAACCATGATGTAATGAAGCTTTACAGTATGACCTCGTAGTGCA 1190
 Qy 134 -----TleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGly 149
 Db 1191 CGCATTATAGTAATCTGGGAGGTATGCTGCCGAGCTTGGAACCTTTTACITTCAGCATTA 1250
 Qy 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp 169
 Db 1251 CATTTAAATGATATCTTCGCCGCGTAGT----- 1277
 Qy 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
 Db 1278 ---CATATTGCAGATGTAGAAATGCAATTAACGCAAACTTGGG-----GAACGA 1325
 Qy 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp----- 206
 Db 1326 GGTTCAC-----GGTACACACACAGCTAAATCTGTACCTGATTTTGACGAT 1370
 Qy 207 ---ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
 Db 1371 GCAATAAACAATTAGCGATAACCTTTGTTGTTGATGCTGGACGACGTTTAACTGTTTCGC 1430
 Qy 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
 Db 1431 CAACCTTCGTTTGAAGAAATACCGTTTCGCTGATAGTACTTTACGTCAGGAATCGGA 1490
 Qy 246 PheGlnProGlyThrProTyrAspLeuLeuLeuAspPheGlnGlnAlaLeuGlu 265
 Db 1491 CAACAAGAGGAACCTTGGTATAATTCACAAATAGTTGAGTTAGGAAAAATTCGCTTAGAT 1550
 Qy 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly--- 284
 Db 1551 CGTACAGGTTTCTTCGAA-----ACAGTTGAAAACCGCAATGATCCTATCATCAATGGTAGC 1604
 Qy 285 ---AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
 Db 1605 AATGATGAGTGGATGCTGATATATAAGTCAACAGCTAACACCGGTAGTAGTATCACTTT 1664
 Qy 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsn 323
 Db 1665 GCTATTGGTTACGCTACAGAGAGTGGTATCATGTTATCAACAAGCTATTAAACAAGATAAT 1724
 Qy 324 LeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThr 343
 Db 1725 TTCTTGGGAACAGGGCGGCGAGTAGTAAAGTATAGCTGCTACGAAAAATGATTATGGTACGAGT 1784

Qy 344 LeuAlaAlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsn 357
 Db 1785 GTCAATTTGGTATTATACCGAACCTTATTTTACTAAAGATGGTAAAGCTCTGGTGGAAT 1844
 Qy 358 TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg 376
 Db 1845 ATTTTCTTTGAAAACTACGATAACTCTAAAGTGATACACTCTCTAACTAATACCGTACG 1904
 Qy 377 AlaPheSerGlyGlyIle-----TyrTyrValArgAspArgAlaGlyIleAspAla 393
 Db 1905 ACTTATGGAAGTAATGTTACTTTAGGTTTCCCTGTAATAAGAAATAACTCTCTATTATGTA 1964
 Qy 394 ArgLeuGly-----AlaGluPheLeuAlaGlu----- 402
 Db 1965 GGATTAGGCCATACCTATATAATAAATAGTAACTTCTCTAGAAATAACCGTAATTTA 2024
 Qy 403 -----GlyArgLysIleProGlySerAspIleAspLeuGly 414
 Db 2025 TATATTCAATCAATGAAATTTAAAGGTAATGCGATTTAAACAAATGACTTTGAT----- 2078
 Qy 415 AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
 Db 2079 -----TTTTCTTTGGTGGAACTATAACAGCCTTAATAGAGGCTAT 2120
 Qy 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
 Db 2121 TTCCCACTAAAGGG-----GTTAAAGCAAGTCTTGGTGGACGAGTACTATTCCAGGT 2174
 Qy 455 SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsn 474
 Db 2175 TCTGATAACAACTACTACAACTAAAGTGCAGATGTACAGGGTTTCTACCCATTAGACAGA 2234
 Qy 475 LysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsn 494
 Db 2235 GATCACCGCTGGTGTGATCTGCAAAAGCATCTGCAGGATATGCAATGTTTGGAAAC 2294
 Qy 495 AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyr 514
 Db 2295 AAGCGTTACCGTCTCTATCAAACTTATACAGCGGGTGGCATTGTTTCATTCACGGGTTT 2354
 Qy 515 GluLeuAspSerIleGlyLeuAlaGlyProAsn----- 535
 Db 2355 GCTTATGGTAGTATT-----GGCCTTAATGCAATTTATGCCGAACATGTTAATGGT 2405
 Qy 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
 Db 2406 ACTTTTAAAGATAAGTTCTGTATGTTGTTGGTAAATGCAATCACAACCTGGAGTGCA 2465
 Qy 540 GluTyrGlnLeuPro-----PheThrArgThr--- 548
 Db 2466 GAACCTTATGTACCACTCCATTTGTGAGTGATAAAGCCAAATAACAGTCCGACCTCC 2525
 Qy 549 -----Leu 549
 Db 2526 CTATTTGTTGATCGCGCAAGTGTGTTGAACTACTAAATGGAATAACAGATAAAATGGATTA 2585
 Qy 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
 Db 2586 GAGAGCAAGTCTTTGAAAGACTTACCTGATTTATGGC-----AAATCAAGCCGTATT 2636
 Qy 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
 Db 2637 CGCGCTCTACAGGTGTCGGATCCCAATGCCAATCTCTCTATTGGACCATTTGTTTCT 2696
 Qy 590 IleAla-----TyrGlyHisSerAspLysLysIleArgTrpHisIleSer 604
 Db 2697 TATGCTAAACCAATTAATAAATGAAATGATGATCGAA---CAGTTCCCAATTTAGT 2753
 Qy 605 LeuGlyThrArgPhe 609
 Db 2754 ATGGGGGCTCTTTC 2768

RESULT 2

US-09-135-166-9
: Sequence 9, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2989 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 390..2768
US-09-135-166-9

Alignment Scores:
Pred. No.: 6.28e-13 Length: 2989
Score: 213.50 Matches: 127
Percent Similarity: 35.04% Conservative: 85
Best Local Similarity: 20.99% Mismatches: 254
Query Match: 6.73% Indels: 139
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-09-135-166-9 (1-2989)

QY	150	AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp	169
Db	1251	CAITTAATGATACCTTCCGCCGTAGT-----	1277
QY	170	PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys	189
Db	1278	---GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTGGG-----GAACGA	1325
QY	190	GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----	206
Db	1326	GGTTAC-----GGTAACAACAACAGTAGTAATTTCTGTACCTGATTTTGACGAT	1370
QY	207	---ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly	225
Db	1371	GCAATAAAACATTAGCGATAACCTTTGTTGTTGATCTGCGACGACGCTTAACTGTTGCG	1430
QY	226	AspPheGluIleThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
Db	1431	CAACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGTACTTTACGTCAGGAAATCGA	1490
QY	246	PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu	265
Db	1491	CAACAAGAAGAACTTGGTATAATTACAAATTAGTTGAGTTAGCAAAATTCGCTTAGAT	1550
QY	266	GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly---	284
Db	1551	CGTACAGTTTCTCGNA-----ACAGTTGAAACCGAATTGATCTATCAATGGTAGC	1604
QY	285	---AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr	303
Db	1605	AATGATGAAGTGGATGCTGATATAAAGTCAAAAGCTAAACACGCGGTAGTATCAACTTT	1664
QY	304	GlyIleArgLeuAspSerGluTyrGlyLeuGlyClyLysIleAlaTyrAspTyrAsn	323
Db	1665	GGTATTGGTTACGGTACAGAGGTGGTATCAGTTATCAATGATCAACAGATTAATAA	1724
QY	324	LeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThrThr	343
Db	1725	TTCTTGGGAACAGGGCGGCGAGTANGTATAGCTGGTACGAAATGATTTGGTACAGT	1784
QY	344	LeuAlaAlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsn	357
Db	1785	GTCAATTTGGTTATACCGAACCCCTATTTACTAAAGATGGTGAAGCTTTGGTGAAT	1844
QY	358	TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg	376
Db	1845	ATTTCTTTGAAACTACGATAACTCTAAAAGTGATACATCCCTCTAACTATAAGCGTACG	1904
QY	377	AlaPheSerGlyGlyIle-----TyrTyrValArgAspArgAlaGlyIleAspAla	393
Db	1905	ACTATGGAAGTAAATGTACTTTAGTTTCCCTGTAATGAAATATACTCCTATTATGTA	1964
QY	394	ArgLeuGly-----AlaGluPheLeuAlaGlu-----	402
Db	1965	GGATTAGCCCATACCTATAATAAATTAGTAACCTTCTCTAGAATATAACCGTAATTTA	2024
QY	403	-----GlyArgLysIleProGlySerAspIleAspLeuGly	414
Db	2025	TATATTCAATCAATGAAATTTAAAGGTAATGGCATTAAACAAATGACTTTGAT-----	2078
QY	415	AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu	434
Db	2079	-----TTTTCTTTGGTGGAACTATATACACCCCTTAATAGAGGCTAT	2120
QY	435	HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu	454
Db	2121	TTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGGTGGACGAGTACTATTCCAGGT	2174
QY	455	SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsn	474
Db	2175	TCTGATAACAAATACTACAAACTAAGTCGATGATACAGGGTTTCTACCCATTAGACAGA	2234


```
QY 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
Db 1277 ATTCAGATCTAGAAATGCAATTAAGCAAACTTGGG-----GAACGAGGTTC 1327
QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
Db 1328 -----GGTAAACAACAGTAGTAATTTCTGCTACCTGATTTTGTACGATGCCAAAT 1372
QY 208 AlaThrValAspLeuAsnValAlaAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 1373 AAACATTACCGTAACCTTTGTTGATGCTGGAGCAGCTTTAACTGTTACCAACTT 1432
QY 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 1433 CGCTTTGAAGGAATAACCGTTTCTGCTGATAGTACTTTACGTCAGGAATGCCCAACA 1492
QY 248 ProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGlnAsn 267
Db 1493 GAAGGAACCTTGGTATATTCACAAATAGTTGAGTTAGGAAATAATTCGCTTAGATCGTACA 1552
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
Db 1553 GGTTCCTCGAA-----ACAGTTGAACCGGAATGATGCTCTCAATGGTAGCAATGAT 1606
QY 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
Db 1607 GAAGTGCATGCTGATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGGTATT 1666
QY 306 ArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrAsnLeuPhe 325
Db 1667 GGTACGGTACAGAGGTGATATTAGTTATCAAGCAAGTGTCAACAAAGATAAATTTCTTG 1726
QY 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAla 345
Db 1727 GGACAGGGCGCGAGTAAGTATAGCTGGTACGAAATGATGATGTCAGAGGTGCTCAAT 1786
QY 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTrp 359
Db 1787 TTGGTTATACCGAGCCCTATTTTACTAAAGATGGTCTAGTCTTGTGCAATGTTTTC 1846
QY 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArgAlaPhe 378
Db 1847 TTTGAAACACTACGATAACTCTAAAGGTGATACATCTCTCACTATAAGCGCTAGCACTTAT 1906
QY 379 SerGlyGlyIle-----TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
Db 1907 GGAAGTAAATGTTACTTTAGTTTCCCTGTAATGAATAAATCACTCTATTATGTAGGATTA 1966
QY 396 Gly-----AlaGluPheLeuAlaGlu----- 402
Db 1967 GGCCATACCTATAATAAAATAGTAACCTTTGCTCTAGAAATAAACCGTAATTTATATATT 2026
QY 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 2027 CAATCAATGAAATTTAAAGGTAATGGCATTAACAACTAACTTTGAT----- 2074
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 2075 -----TTTCTTTTGGTTGGAAGTAATAACAGCCCTTAATACAGGCTATTTCCTCA 2122
QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 2123 ACTAAAGGG-----GTTAAAGCAAGTCTTGGTGGAGGATTACAATTCACGGTTCTGTAT 2176
QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 2177 AACAAATACCTACAACTAAGTACAGATGTACAGGTTTCTACCCATTTAGACAGATCAC 2236
QY 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAsp 496
Db 2237 CTCTGGTGTCTATCTGCAAAAGCATCTGCAGGATATGCAATGGTTTGGTGAACACCGGT 2296
QY 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
```

```
Db 2297 TTACCGTTCATCAAACTTATACAGCGGTGGCATTTGTTTCATTCACCGGTTTTCCTTAT 2356
QY 517 AspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2357 GGTAGCAAT-----GGGCCTAACGCAATTTATCAAGGTCAAAATAAATAAATTTAAT 2407
QY 526 -----GlySerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGln 542
Db 2408 AAGATAAGTTCTGATGTGTTGGTGAATGCAATCCCTACAGCTAGCGCAGAGTTAATT 2467
QY 543 LeuPro-----PheThrArgThr----- 548
Db 2468 GTGCCAACTCCATTGTGTGAGTGATAGAGAGTCAAAATACAGTCCGCAACTCCCTATTGTT 2527
QY 549 -----LeuSerGlyAla 552
Db 2528 GATCGCGCAAGTGTGTTGGAATCTAAATGAAATACAGATAAAATGGATTAGAGAGCAAT 2587
QY 553 ValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGly 572
Db 2588 GTCTTGAAGACATTACCGGATTATGGC-----AAATCAACCGCTACTCGCGCCTCT 2638
QY 573 SerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla--- 591
Db 2639 ACAGGTGTCGATTCCAATGCAATCTCTTAGTGACCACTGGTATTCTTCTATGCTAAA 2698
QY 592 -----TyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThr 607
Db 2699 CCAATTAATAAATATGAAATGATGTCGAA---CAGTTCCAATTTAGTATTGGGGGT 2755
QY 608 ArgPhe 609
Db 2756 TCATTTC 2761

RESULT 5
US-09-135-166-7
; Sequence 7, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
```

RESULT 6
US-08-942-046-7
; Sequence 7, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele

APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto

STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/942,046
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522

FILING DATE: 12-SEP-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I

REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MTS:jb

TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2974 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 386..2761

us-08-942-046-7

Alignment Scores:
Pred. No.: 1.32e-12 Length: 2974
Score: 210.50 Matches: 121
Percent Similarity: 34.88% Conservative: 75
Best Local Similarity: 21.53% Mismatches: 241
Query Match: 6.63% Indels: 125
Gaps: 21

US-09-857-669-2 (1-609) x US-08-942-046-7 (1-2974)

Qy 134 IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu 151
Db 1193 ATAGGTAATCTGGGAGGTATCTGCCGAGCTTGAACCTTTACTTTTCAGCATTCATTTA 1252
Qy 152 AlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAsp 171
Db 1253 AATGATACTTTCCGCGGTAGT-----GAT 1276
Qy 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
Db 1277 ATTCAGCATGTAGAAATGCAATTAAGCAAACTTGGG-----GAACGAGGTAC 1327
Qy 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
Db 1328 -----GGTAACACACACAGATAAATCTGTACCTGATTTTGACGATGCAAAAT 1372
Qy 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227

Db 1373 AAAACATTAGCGATAACCTTTGTTGTTGATGCTGCGAGCAGCTTTAACTGTTTCACCAACTT 1432
Qy 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 1433 CGCTTTGAAGGAATAACCGTTTCGCTGATAGTAGTACTTTCAGTCAGGAAATGCGCAACA 1492
Qy 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
Db 1493 GAAGCAACTGGTATATTCACAATTAGTTGAGTTAGGAAAAATTCGCTTAGATCGTACA 1552
Qy 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
Db 1553 GGTTCCTCGAA-----ACAGTTGAAACCCGAATTGATCTCATCAATGGTAGCAATGAT 1606
Qy 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
Db 1607 GAAGTGGATGCTGATATATAAGTCAAGACGTAACACGGGTAGTATCAACTTTGGTATT 1666
Qy 306 ArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
Db 1667 GGTTCAGGTACAGAGAGTGGTATTAGTTATCAACCAAGTGTCAACCAAGATAATTCTTG 1726
Qy 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAla 345
Db 1727 GGAACAGGGCGGCGAGTAAAGTATAGTGTGTAAGAAAAATGATTATGGTACGAGTCAAT 1786
Qy 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTrp 359
Db 1787 TTGGTTATACCGGACCTATTTTACTAAGATGGTGTAAAGTCTGGTGGAAATGTTTTC 1846
Qy 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArgAlaPhe 378
Db 1847 TTTGAAACTACGATAACTCTAAAGTGATACATCCTCTAACTATAGCGTACCACTTAT 1906
Qy 379 SerGlyGlyIle-----TyrTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
Db 1907 GGAAGTAATGTTACTTTCCTGCTTAATGAAATAACTCTCTATATGTTAGGATTA 1966
Qy 396 Gly-----AlaGluPheLeuAlaGlu----- 402
Db 1967 GCCATACCTATAATAAATTTAGTAACTTTGCTCTAGAATAATAACCGTAATTTATATATT 2026
Qy 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 2027 CAATCAATGAATTTAAAGGTAATGGCATTTAAACAAATGACTTTGAT----- 2074
Qy 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
Db 2075 -----TTTTCTTTTGGTGGAACTATAACAGCCCTTAATAGAGCGTATTTCCCA 2122
Qy 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 2123 ACTAAAGGG-----GTTAAAGCAAGTCTTGTGCGACAGTTACAAATCCAGGTTCTGAT 2176
Qy 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPheThrProGluAsnLysLys 476
Db 2177 AACAAATACTACAAACTAAGTGCAGATGTACAGGGTTTCTACCCCTTAGACAGAGATCAC 2236
Qy 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAlaAsp 496
Db 2237 CTCCTGGGTGTATCTGCAAAAGCATCTGACGATATGCAATGCTTTTGGAAACAAGCGT 2296
Qy 497 ValProSerGlyLeuMetPheArgSerGlyAlaSerValArgGlyTyrGluLeu 516
Db 2297 TTACCGTTCTATCAAACTTATACAGCGGGTGGCATTTGGTTCATTACGCGGTTTTGCTTAT 2356
Qy 517 AspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2357 GGTAGCAT-----GGGCTTAACCGCAATTTATCAAGCTCAAAATAAATTAAT 2407
Qy 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542

Db 2408 AAGATAAGTTCTGATGTGATTGGTGGTAATGCAATCCCTACAGCTAGCGCAGAGTTAATT 2467
QY 543 LeuPro-----PheThrArgThr----- 548
:::|||||
Db 2468 GTGCCAACTCCATTGTGAGTGATAAGAGTCAAAATACAGTCCGAACCTCCCTATTGTT 2527
QY 549 -----LeuSerGlyAla 552
:::|||||
Db 2528 GATCGCGCAAGTGTGGAACTACTAAATGGAATFCAGATAAAATAGGATTAGAGAGCAAT 2587
QY 553 ValPheHisAspMetGlyAspAlaAlaAlaAsnPhelYsArgMetLysLeuLysHisGly 572
||| |:::|
Db 2588 GTCTTGAAGACATTACCGGATATGGC-----AAATCAAGCGGTACTCGCGCTCT 2638
QY 573 SerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla--- 591
:::|||||:::|
Db 2639 ACAGGTGCGGATCCCAATGCGCATCTCTAGTGGACCACTGGTATTTCTTATGCTAAA 2698
QY 592 -----TyrGlyHisSerAspLysIleArgTrpHisIleSerLeuGlyThr 607
||| |:::|
Db 2699 CCAATTAAAAAATGAAATGATGATGCGAA---CAGTTCCAATTTAGTATGGGGGT 2755
QY 608 ArgPhe 609
|||
Db 2756 TCTTTC 2761
RESULT 7
US-08-433-522A-1
; Sequence 1, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jfb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2949 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 75..2465
US-08-433-522A-1

Alignment Scores:
Pred. No.: 1 67e-12 Length: 2949
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24
US-09-857-669-2 (1-609) x US-08-433-522A-1 (1-2949)
QY 61 AspSerGluLeuLysAspMetValGluGluHisLeuProLeuLysLeuThrGlnGlnGlu 80
:::|:::|
Db 624 GAATCTCTTAGTACAGCATACATACAAAGACAAATGCAATACACCTGATCTCTGCTGG 683
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
:::|:::|
Db 684 AAATTATGGGAAATAAATTTGAAGTGCAGCAATTCGAGAAAGATTTGCAGTCAATTCGT 743
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
:::|:::|
Db 744 GATTATTATTAAATAATGGCTATGCCAAAGCACAAATTTACTAAACGGATGTTTCAGCTA 803
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
||| |:::|
Db 804 AATGATCAAAAAACAAAGTTAATGTAAACCATTCATGTAATGAAGCTTTACAGTATGAC 863
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
||| |:::|
Db 864 CTTCTGATGTCACGCATTTATAGGTAATCTGGAGGTATGTCGCCGAGCTTGAACCTTTA 923
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
||| |:::|
Db 924 CTTTCACCATACATTTAATGATGACTTTCGCCGCTAGT----- 962
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
||| |:::|
Db 963 -----GATATTGCAGATGTAGAAATGCAATTAAGCAAAACTTGA--- 1004
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
:::|:::|
Db 1005 -----GAACGCGGTTACGGTAGCGCAACGGTAAATTCAGTA-----CCT 1043
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
||| |:::|
Db 1044 GATTTTGTATGATGCAATAAATACATACCGTATACCTCTGTTGTGATGCTGGACGAGCT 1103
QY 222 IleAlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSer 241
:::|
Db 1104 TTAACCTGTCGCCAACTTCGCTTTTGAAGAAATACCGCTTCTGCTGATAGCACCTTACGT 1163
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
:::|
Db 1164 CAGAAATGCGCCCAACAAAGAACTTGGTATATAATTCACAAATTAGTTAGTATAGGAAA 1223
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
||| |:::|
Db 1224 ATTCGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAACCGAATTTGATCCT 1277
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
:::|
Db 1278 ATCAATGCTAGTAAATGATGAAGTGGATCTCGTATATAAAGTCAAAAGACGTAACACGGGT 1337
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
:::|
Db 1338 AGTATCAACTTTGGTATTGTTACGGTACAGAGAGTGGTATTAGTTATCAACAGCAAGTGT 1397
QY 320 AspTyrTrpAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
||| |
Db 1398 AACCAAGATAATTCTTGGGAACAGGGCGGCACATAAGATATAGCTGGTACGAAATATGAT 1457
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
||| |:::|

```
Db 1458 TATGTCAGCAGTGTCAATTTGGGTTATACCGAGCCCTATTTTACTAAAGATGGTGTAAGT 1517
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1518 CTGGTGGAAATGTTCTTTTGAAACACTACGATACTCTAAGAGTGATACATCTCTTAAC 1577
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1578 TATAAGCGCTACGACTTACGGAAGTAATGTTACTTCTAGGTTTCCCTGTAAATGAAATAAC 1637
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1638 TCCTATTATGTAGGATTAGGTCATACCTATATAAATGATTAACCTTTCCTCTAGATAT 1697
Qy 403 -----GlyArgLysIleProGlySerAsp 410
Db 1698 AACCGTAATTTATATTCATCAATCAATGAATTTAAGGTATGTCATTAACAACAATGAC 1757
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTCTTTTGGTTGGAACTATAACAGCCTT 1793
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 AATAGAGCGCTATTTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGGTGGACGAGTT 1847
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCAGGCTTCGATAACAATACTACAACACTAAGTGCAGATGTACAGCGTTTCTAC 1907
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAA 1967
Qy 491 AlaArgAspAsnAlaAspValProSerClyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTGTTTGGAAACAGCGTTTACCCTTCTATCAACTTATACAGCGGGTGGCATCGGTTC 2027
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGTTTGTCTTATGTGTAGTATT-----GGACCTAACGCATTTATGCCCA 2078
Qy 526 -----GlySerValLeuProGlu 531
Db 2079 TATGGTAATGCTAGTGTACTGCTACTTTTAAAGAGATAAGTCTGTGATGTGATGGTGGT 2138
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCGCAGAGTTAATGTGTGCCAATCCATCTTGTGAGCGATAAG 2198
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2199 AGCCAAATACGGTCCGAACCTCTTATTTCTTCATCGCGCAAGTGTGTGAATACTAAA 2258
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGATGATAGAGCGATGCTATTAAAGATTTGCCTGATTATGGC 2318
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AATCAACCGGTATTCGGCCCTCTACAGGTGTCGGATTCCTCAATGGCAATCTCTTTGGG 2378
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGCTATTCTTATGCCAACCAATTAATAAATATGAATATGATGATGTCGA--- 2435
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCAAATTTAGTATTGGAGGTTCTTTC 2465
RESULT 8
US-09-135-166-1
; Sequence 1, Application US/09135166

; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2949 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 75..2465
; US-09-135-166-1

Alignment Scores:
Pred. No.: 1,67e-12 Length: 2949
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-09-135-166-1 (1-2949)
Qy 61 AspSerGluLeuLysAspMetValGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 624 GAATCTGTTAGTAGCAGTACATTACAAGACAAATGGAATTAACAACCTGATTCTTGGTGG 683
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 684 AAATTATGGGAAATAAATTTGAAGGTCCGCCAATTCGAGAAAGATTTGCAGTCAATTCGT 743
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 744 GATTATTATTAAATAATGCGTATGCCAACACAAATTAATAACGGATGTTTCAGCTA 803
Qy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
|||||
|||||
|||||
```

Db 804 AATGATGAAAAACAAAGTTAAATGTAACCAATGATGTAATAAGAGCTTTACAGTATGAC 863
Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 864 CTTGCTAGTGCACCAATATAGTAATCTGGAGGTATGCTGCGGACCTTGAACCTTTA 923
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 924 CTTTCAGCATTACATTTAAATGATACTTTCGCGCGTAGT----- 962
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 963 -----GATATTTCAGATGTAGAAAATGCAATTAAGCAAAAACCTTGA--- 1004
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1005 -----GNACCGGTAGCGTAGCGCAACGGTAATTCAGTA-----CCT 1043
Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
Db 1044 GATTTTGATGATGCAAAATAAACAATAGCGATAACCCCTTGTGTGATGCTGGACGAGT 1103
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1104 TTAACCTGTCGCCCAACTTCGCTTTCAAGCAAAATACCGCTTCTGCTGATAGCACTTTACGT 1163
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1164 CAGGAATGCCCAACAGAAGAACTTGGTATATATTCACAATTAGTTGAAGTTAGGAANA 1223
Qy 262 GlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1224 ATTCGCTTAGATCGTACAGTTTCTTCGAA-----ACAGTCGAAACCGCAATGATCCT 1277
Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1278 ATCAATGAGTAGATGATGAAGTGGATGCTGATATAAAGTCAAGAACGTAACACGGGT 1337
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1338 AGTATCAACTTTGGTATTTGGTACGGTACAGAGAGTGTATAGTTATCAAGCAAGTGT 1397
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1398 AAACAAGATAATTTCTGGGAACAGGCGCGCAGTAAAGTATAGTGTGTACCAAAAATGAT 1457
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1458 TATGATACGAGTGTCAATTTGGGTATATACCGAGCCCTATTTTACTAAAGATGGTGAAGT 1517
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1518 CTTGGTGAAGATGTTTCTTTGAAACCTACGATACTCTAAAGTGATACCTCTCTAAC 1577
Qy 373 LeuGluTyrArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1578 TATAAGCCTACGACTTACGGAAGTAATCTTACTTTAGTTTCCCTGTAAATGAAAATAAC 1637
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1638 TCCTATTATGAGGATTAGGTATAGCTATATAAATAATAGTAACCTTGTCTAGAAATAT 1697
Qy 403 -----GlyArgLysIleProGlySerAsp 410
Db 1698 AACCGTAATTTATATATTCATCAATGAATTTAAAGGTAATGGCATTAATAACAAATGAC 1757
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTCTTTTGGTGGAACTATAACAGCCTT 1793
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 AATAGAGCTATTTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGGTGGACGAGTT 1847

Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCCAGGTTCTGATAACAAATACTACAAATAAGTGCAGATGTACAGGGTTTCTAC 1907
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACCTCTGGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAAAT 1967
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTTTTGGAAACAACGCTTACCGCTTCATCAAACTTATACAGCGGGTGGATCGGTTCA 2027
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGGTTTGTGTTGTTAGTAGTATT-----GGACCTAACCCCAATTTATGCCGAA 2078
Qy 526 -----GlySerValLeuProGlu 531
Db 2079 TATGTAATGCTAGTGGTACTGTTACTTTTAAAGATAAGTTCTGTGATGTGTTGGTGGT 2138
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCGCAGAGTAAATTTGTGCCAACTCCATTGTTGAGCGATAAG 2198
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2199 AGCCAAAATACGGTCCGAACTCCTTATTGTTGATGCGGCAAGTGTGTTGGAATACTAAA 2258
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGGATTAGAGCGCATGTATTAAAAGATTGCTGATTATGGC 2318
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AAATCAAGCCGTATTTCGCGCCTCTACAGGTGTCGAGTTCGAATGCCAATCTCCTATTGGG 2378
Qy 585 PropheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGGTATTCTCTTATGCCAAACCAATTTAAAATAATGAANAATGATGATGTCGAA --- 2435
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCAATTTAGTATTGGAGTTCTTTC 2465

RESULT 9
US-08-942-046-1
; Sequence 1, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942.046
; FILING DATE:

```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 08/433,522
:   FILING DATE: 12-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
:   NAME: STEWART, Michael I
:   REGISTRATION NUMBER: 24,973
:   REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (416) 595-1155
:   TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 1:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 2949 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   FEATURE:
:     NAME/KEY: CDS
:     LOCATION: 75..2465
:   US-08-942-046-1

```

```

Alignment Scores:
Pred. No.: 1,67e-12 Length: 2949
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 4 Gaps: 24

```

US-09-857-669-2 (1-609) x US-08-942-046-1 (1-2949)

```

Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 624 GAATCTGTTAGTAGCAGTACATTAACAGAACAAATAGGAATTACACACCTGATCTTGGTGG 683
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 684 AAATTATGGGGAATAAATTTGAAGGTCGCCAATTCGAGAAAGATTTCGACATCAATTCGT 743
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 744 GATTATTATTAATAATATGCGTATGCCAAAGCACAAATTAATAAACCGATGTTTCAGCTA 803
Qy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 804 AATGATGAAAAACAAAGTTAATGTAACTATGATGTAATGAAAGGTTTACAGTATGAC 863
Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 864 CTTCGTAGTCACGCATTATAGGTAATCTGGAGGTATGTCGCCGACCTTGAACCTTTA 923
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 924 CTTTCAGCATTACATTTAATGATACTTTCGCCCGCTAGT----- 962
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 963 -----GATTTGCGAGATGTAGAAATGCAATTAAGCAAAACCTTGGA--- 1004
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1005 -----GAACGCGTTACGGTAGCGCAACGGTAAATTCAGTA-----CCT 1043
Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1044 GATTTTGATGATGCAAAATAAACATATAGCGATAACCCCTTGTGTGATGCTGGACGAGGT 1103
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1104 TTAAGTGTCCGCCAACCTTCGCTTTGAAGGAATAACCGCTTCTGCTGATAGCACTTTACGT 1163

```

```

Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGln 261
Db 1164 CAGGAAATGCGCCAAAGAAAGAACTTGGTATAATTACCAATTAGTTGACTTAGGAAAA 1223
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1224 ATTCGCTTAGATCGTACAGAGTTTCTTCGAA-----ACAGTCGAAACCGAATTTGATCT 1277
Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1278 ATCAATGGTAGTAATGATGAAGTGTCTGTATATAAAGTCAAGAAGACGTAAACACGGGT 1337
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyr 319
Db 1338 AGTATCAACTTTTGGTATTTGCGGTACAGAGTGTCTATTAGTTATCAAGCAAGTCTT 1397
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1398 AACCAAGATAAATTTCTTGGGAACAGGGCGCAGTAAGTATAGCTGTGTCGAAAAATGAT 1457
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1458 TATGGTAGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTACTAAAGATGGGTAAAGT 1517
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1518 CTTCGTGCAAAATGTTTCTTTGAAAACTACGATAAATCTCTAAAGCTGATACATCTCTAAC 1577
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1578 TATAAGCTACGACTTACGGAAGTAATGTTACTTTTAGCTTTCCCTGTAAATGAAATAAC 1637
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1638 TCCTATTATGTAGGATTAGGTCATACCTATATAAATATTAGTAACTTTGCTCTAGAATAT 1697
Qy 403 -----GlyArgLysIleProGlySerAsp 410
Db 1698 AACCGTAATTTATATATCAATCAATGAATTTAAAGGTAATGGCATTTAAACAAATGAC 1757
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTTCTTTTGGTTGGAACTATAACAGCCCTT 1793
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 AATAGAGCTATTTCCCAACTAAAGGG-----GTTAAAGCAAGTCTTGGTGGAGGAGTT 1847
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCAGGCTTCGTGATAACAATACTACAACTAAGTACAGATGTACAGGGTTTCTTAC 1907
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACCTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAA 1967
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTTTTGGAAACACAGCGTTTACCCTTCTATCAAACTTATACAGCGGGTGGCATCGGTCA 2027
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGGTGTTCGTTATGTTAGTAGTATT-----GGACCTAACGCAATTTATGCCGA 2078
Qy 526 -----GlySerValLeuProGlu 531
Db 2079 TATGGTAATGCTAGTGGTACTGTTTAAAGAAAGATAAGTTCTGATGTGATGGTGGT 2138
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCGCAGAGTAAATTTGGCCCACTCCATTTTCTGAGCGATAAG 2198

```

```
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2199 AGCAAAATACGTCGCAACCTCTTTATTTGTTGATGCGGCAAGTGTGGTAATACTAAA 2258
QY 558 -----GlyAspAlaAlaAsnPhelysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGGATTAGAGAGCGATGATTAAAGATTTGCCGTGATTATGGC 2318
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AAATCAAGCGGTATTCGGCTCTACAGGTGCGGATTCGAATCGCAATCTCTATTGGG 2378
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGGTATCTCTTATGCCAACCAATTAATAATATCAAAATGATGATGTCGAA--- 2435
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCAAATTTAGTATGGAGGTTCTTTC 2465

RESULT 10
US-08-433-522A-5
; Sequence 5, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-08-433-522A-5

Alignment Scores:
Pred. No.: 1,67e-12 Length: 2950
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
```

```
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24
US-09-857-669-2 (1-609) x US-08-433-522A-5 (1-2950)
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 883 GAATCTGTTAGTACGACTACATTACAGAACAATGAATACAAACCTGATCTTGTGGTGG 942
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 943 AAATTATGGGAAATAAATTTGAAGTGCACAATTCAGAGAAGATTTCAGATCAATTCGT 1002
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1003 GATTATTATTAAATATGCGTATGCCAAAGCACAAATTAATAACGGATGTTCACTA 1062
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1063 AATGATCAAAAACAAAGTTAATGTAACTTGAATTAATGAAGGTTTACAGTATGAC 1122
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1123 CTTGCTAGTCACGCATTTATAGTAATCTGGAGGTATGCTGCCGAGCTTGAACCTTTA 1182
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1183 CTTTCAGCATTTAATTAATGATCTTCCGCGGTAGT----- 1221
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1222 -----GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTGA--- 1263
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1264 -----GAACGCGGTTACGGTAGCGCAACGGTAAATTCAGTA-----CCT 1302
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValValValValValValVal 221
Db 1303 GATTTCGTATGATGCAATAAATTAACATTAACGATACCTTTGTTGTTGATGCTGGACG 1362
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1363 TTAACCTGTCGCCAACTTCGCTTTGAAGAAATACCGTTCTGCTGATAGCATCTTACGT 1422
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1423 CAGGAAATGCGCCCAACAAAGAAAGAACTTGGTATATAATTCACAATTTAGTTAGGAAA 1482
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1483 ATTCGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAACCGCAATTTGATCCT 1536
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1537 ATCAATGGTAGTAATGATGAAGTGGATGCTGATATAAAGTCAAAAGACGTAACACGGT 1596
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
Db 1597 AGTATCAACTTTGGTATTGTTACGGTACAGAGAGTGTATTAGTTATCAACGCAAGTGT 1656
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1657 AACAAAGATAATTTCTTGGGAAACAGGGCGGCACTAGTATAGTGGTACGAAATATGAT 1716
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1717 TATGCTAGGAGTGTCAATTTGGGTTATACCGGAGCCCTATTTTACTAAAGATGGTGAAGT 1776
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1777 CTTGGTGAATGTTTCTTTTGAAGAACTACGATAACTCAAAAGCTGATACATCCTCTAAC 1836
```

```
Oy 373 LeuGluLysArgAlaPheSerGlyGlyLe-----TrpTyrValArgAspArgAla 389
Db 1837 TATAAGCGTAGCACTACGGAAGTAAGTCTTACTTTAGCTTTCCCTGTAATGAAATCAATAC 1896
Oy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1897 TCCTATTATGATGAGGATTAGGTACCTATATATAATTAAGTAACTTTGCTCTAGAAAT 1956
Oy 403 -----GlyArgLysIleProGlySerAsp 410
Db 1957 AACCGTAATTTATATTCATCAATCAATGAATTAANGGTATGCAATTAACAACAAATGAC 2016
Oy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2017 TTTGAT-----TTTTCTTTTGGTTGGAACTATAACAGCCTT 2052
Oy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2053 AATAGAGGCTATTCCCAACTAAAGG-----GTTAAGCAAGTCTTGGTGGAGCGATT 2106
Oy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2107 ACTATTCAGGCTCTGATACAAATACTACAACATAAGTCGAGATGTACAGGTTTCTAC 2166
Oy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 2167 CCATTAGACAGAGATCACTCTGGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAT 2226
Oy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyValaSerSer 510
Db 2227 GGTGTTGGAAACAACGGTTTACCCTTCTATCAAACTTATACAGCGGGTGGCATCGGTTC 2286
Oy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2287 TTACGTGCTTTGCTTATGCTAGTATT-----GCACCTAACGCCAATTTATCCCGAA 2337
Oy 526 -----GlySerValLeuProGlu 531
Db 2338 TATGCTAATGCTAGTGTACTGCTACTTTTAAGAAAGATAAGTCTGTGATGTGTTGGTGGT 2397
Oy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2398 ANTGCATCGCTACAGCTAGCGCAGAGTTAATGTGCCAACTCCATTTGTGAGCGATAAG 2457
Oy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2458 AGCAAAATACGGTCCGNAACCTCTTATTGTTGATCGGCAAGTGTGGAATACTAAA 2517
Oy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2518 TGGAAATCAGATAAAATGGATTAGAGAGCGATGTATTAAAGAGATTGCCCTGATTATGGC 2577
Oy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2578 AATCAACCGCTATTTCGGCCCTCTACAGGTGTCGGATTCCCAATGGCAATCTCCTATTGGG 2637
Oy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2638 CCATTGGTATTCTTATGCCCAACCAATTAANAATATGAATGATGATGTCGAA--- 2694
Oy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2695 CAGTTCCAAATTTAGTATTGGAGGTTCTTTTC 2724

RESULT 11
US-09-135-166-5
; Sequence 5, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
```

```
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-09-135-166-5

Alignment Scores:
Pred. No.: 1,67e-12 Length: 2950
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-09-135-166-5 (1-2950)
Oy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 883 GAATCTGTTAGTAGCAGTACATACAAAGCAAAATGGAATTAACAACCTGATTCTTGGTGG 942
Oy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 943 AATATTATGGGAAATAAATTTGAAGGTGCGCAATTCGAGAAGAGATTTGACGTCAATTCGT 1002
Oy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1003 GATTATTATTAAATAATGGCTATGCCAAACACCAAAATTAACACGGATGTTTCAGCTA 1062
Oy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1063 AATGATGAAAAACAAAAGTTAATGTAACCATTTGATGTAATGAAGTTTACAGTATGAC 1122
Oy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1123 CTTCGTAGTCACCGCATATTAGTAACTCTGGAGAGGTATGTCTCGCGACCTTTGAACCTTTA 1182
```

```
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
Db 1183 CTTTCAGCATTACATTTAAATGATACTTCCGCGTAGT----- 1221
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1222 -----GATATTCAGATGTAGAAATGCAATTAAGCAAAACTTGA--- 1263
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1264 -----GAACGCGTTCAGGTACGCGCAACGGTAAATTCAGTA-----CCT 1302
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
Db 1303 GATTTTGTATGATGCAAAATAAACAATAGCGATAACCCCTTGTGTGATGCTGGACGCGT 1362
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1363 TTAACGTGTCGCCAACTTCGCTTGAAGGAATAACCGTTTCTGCTGATAGCACCTTTACGT 1422
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1423 CAGGAATCGGCCAACAAAGAACTTGGTATAATTCACAATTAGTTCAGTTAGGAAAA 1482
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1483 ATTCGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAACCGAATTGATCCT 1536
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1537 ATCAATGGTGGTAATGATGAAGTGGATGCGTATATAAATGTCNAAAGAACTAACACGGGT 1596
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1597 AGTATCAACTTGGTATTGTTACGGTACAGAGGTGGTATTAGTTATCAACGAAGTGT 1656
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1657 AAACAAGATAATTTCTTGGAAACAGGCGGCGAGTAAAGTATAGCTGGTACGAAATAATGAT 1716
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1717 TATGGTACAGGTGTCAATTTGGGTTATACCGGACCCCTATTTACTAAAGATGGTGTAAAT 1776
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1777 CTGGTGGAAATGTTTCTTTGAAACTACGATAACTCTAAAGTATGATCACTCCTCTAAC 1836
QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1837 TATAACGCTACGACTTACGGAAGTAAATGTTACTTTTAGTTTCCCTGCTAAATGAAATAAC 1896
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1897 TCCTATTATAGGATTAGGTATGATACCTATAATAAATTAGTAACCTTGTCTAGAAATAT 1956
QY 403 -----GlyArgLysIleProGlySerAsp 410
Db 1957 AACCGTAATTTATATTCATCAATCAATGAAATTTAAAGGTAATGGCAATTAACAAATGAC 2016
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2017 TTTGAT-----TTTTCTTTTGGTGGAACTATAACACCCCT 2052
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrLeu 450
Db 2053 AATAGAGCGTATTTCCTCAACTAAAGG-----GTTAAAGCAAGTCTGTGTGACGAGTT 2106
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2107 ACTATTCAGGTCTCGATACAAATAACTACAACTAAGTCAGATGTACAGGTTTCTAC 2166
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
```

```
Db 2167 CCATTAGACAGAGATCACCTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAAAT 2226
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyAlaSerSer 510
Db 2227 GGTTTTGGAAACAAGGTTTACCGTCTCTATCAAACTTATACAGCGGGTGGCATCGGTTCA 2286
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2287 TTACGTGGTTTTCCTTATGTTAGTATT-----GGACCTAACCGCAATTTATGCCGAA 2337
QY 526 -----GlySerValLeuProGlu 531
Db 2338 TATGTTAATGTTAGTGTACTGCTACTTTTAAAGAGATAAGTTCTGATGTGTTGTTG 2397
QY 532 ArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2398 AATGCAATCGCTACAGCTAGCGCAGAGTAAATTTGTGCCAACCTCATTGTGAGCGATAAG 2457
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2458 AGCCAAATACGGTCCGAACCTCCTTATTGTTGATCGCGCAAGTGTGGAATACTAA 2517
QY 558 -----GlyAspAlaAlaAlaAsnPhelLysArgMet----- 567
Db 2518 TGGAAATCAGATAAAATGGATTAGAGAGCGATGTATTAATAAAGATTGCTGATTATGGC 2577
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2578 AATCAAGCGGTATTCGCCCTCTACAGGTGCGGATTCCAATGGCAATCTCCTATTGGG 2637
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2638 CCATTGCTATTCTTATCCCAACCAATTAATAAATATGAAATGATGATGTCGAA--- 2694
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2695 CAGTCCCAATTTAGTATTGGAGGTTCTTTC 2724
```

RESULT 12

```
US-08-942-046-5
; Sequence 5, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2950 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 334...2724

US-08-942-046-5

Alignment Scores:

Pred. No.: 1,67e-12 Length: 2950
Score: 209.50 Matches: 130
Percent Similarity: 35.8% Conservative: 103
Best local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 4 Gaps: 24

US-09-857-669-2 (1-609) x US-08-942-046-5 (1-2950)

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
DB 883 GAATCTGTTAGTAGCAGTACATTACAGAACAAATAAGGAATTACAACTGATCTTGGTGG 942
QY 81 GluValLeuAspLysGluGlnThrPheLeuAlaGluAlaProAspAsnValLys 100
DB 943 AAATATGGGGAATAATTTGAAGTGCCCAATTTCGAGAAAGATTGCGAGTCAATTCGT 1002
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
DB 1003 GATTATTATTAAATAATGCTATGCCAAGACAAATAATTACTAAACGGATGTTACGCTA 1062
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGluProArgThrLys 133
DB 1063 AATGATGAAACAAACAAAGTAATGTAACCATTTGATGTAATGAAGGTTTACAGTAGTAC 1122
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
DB 1123 CTTTCGTAGTCCAGCATTTATAGTATATCTGGAGGTATGCTGCCGAGCTTGAACCTTTA 1182
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
DB 1183 CTTTCAGCATTTAATATGATCTTTCCGCCGTAGT----- 1221
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
DB 1222 -----GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTGGAA--- 1263
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGluValAsnThrArgAlaAlaValAsnPro 205
DB 1264 -----GAAGCCGGTTACGTAGCCGCAACGCTAAATTCAGTA-----CCT 1302
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
DB 1303 GATTTTGTATGATGCAATAAATAACATTAGCGATAACCCCTTCTGTTGATGTCGACGACGT 1362
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
DB 1363 TTAACCTGTCGCAACTTCGCTTTGAAGAAATACCGCTTCTGCTGATAGCACTTTTACGT 1422
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
DB 1423 CAGGAAATGCGCCCAACAAAGAGGAACTTGGTATATAATTCACAATTAGTTAGTAGGAAA 1482
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281

DB 1483 ATTCGCTTAGATCGTACAGGTTTCTTCGAA-----ACAGTCCAAAACCGAATTGATCCT 1536
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
DB 1537 ATCAATGGTAGTAATGATGAAGTGGATGCGTATATAAGTCAAGACGATACACGGGT 1596
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
DB 1597 AGTATCAACTTTGGTATTGGTTACGGTACAGAGTGGTATTAGTTATCAACCAAGTGT 1656
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
DB 1657 AAACAAGATAAATTTCTTGGACAGCGGCGCAGTAGTAGTATAGTACGAAAAATGAT 1716
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
DB 1717 TATGGTACGAGTGTCAATTTGGGTTATACCGGCCCTATTTTACTAAAGATGGTGAATG 1776
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
DB 1777 CTTGGTGGAAATGTTCTTTTGAAGAACTACGATAACTCTAAAGTGTATACCTCTAAC 1836
QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
DB 1837 TATAAGCTACGACTTACGGAAGTAATGTACTTCTAGTTTCCCTCTAAATGAAAAATAC 1896
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
DB 1897 TCCTATTATGTAGGATTAGGTACATCTATATAAATAATTAGTAACCTTCTCTAGAATAT 1956
QY 403 -----GlyArgLysIleProGlySerAsp 410
DB 1957 AACCGTAATTTATATATCAATCAATGAAATTTAAAGGTAATGGCATTAACAAATGAC 2016
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
DB 2017 TTTGAT-----TTTCTTCTTTGGTTGGAACATAACAGCCCTT 2052
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
DB 2053 AATAGAGGCTATTTCCCACTAAAGG-----GTTAAGCAAGTCTTGGTGGCAGGATT 2106
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
DB 2107 ACTATTCAGGTTCTGATAACAAATACTACAACTAAGTGCACATGTACAGGCTTTCTAC 2166
-QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
DB 2167 CCATTAGACAGAGATCACCTCTCGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAT 2226
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
DB 2227 GGTTTTGGAAACAGCGTTTACCGTCTATCAAACTTATACACGGGTGCGCTGCTCA 2286
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
DB 2287 TTACGTGGTTTCTGTTATGTGTAATTT-----GGACCTAACGCAATTTATGCGAA 2337
QY 526 -----GlySerValLeuProGlu 531
DB 2338 TATGGTAATGGTAGTGTACTGTTTAAAGAGATAAGTTCTGATGCTGATTTGGTGGT 2397
QY 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
DB 2398 AATGCAATCCCTACAGTACGCGAGAGTTAAATTTGTGCCAACTCCATTTGTGACGGATAAG 2457
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
DB 2458 AGCAAAATACGCTCCGAACCTCTTATTTGTTGTCATGCGGCAAGTGTGGAATACTAAA 2517
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
DB 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567

Db 2518 TGAATACAGATAAAATGGATTAGACAGCGATGTATTAAAAAGATGCTGCTGATTATGCG 2577
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAla 584
Db 2578 AATCAAGCGGTATTCGCCCTCTACAGGTGCGGATTCGAATCCATGGCAATCTCTATTGGG 2637
QY 585 ProPheSerPheAspIleala-----TyrGlyHisSerAspLysLysIle 599
Db 2638 CCATTGGTATTCTCTTATGCCAACCAATTAATAAATATGAAATATGATGATGCGAA--- 2694
QY 600 ArgTyrPheHisSerLeuGlyThrArgPhe 609
Db 2695 CAGTTCCAATTAGTATTGGAGGTTCTTTC 2724

RESULT 13

US-08-433-522A-3
; Sequence 3, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..2764
US-08-433-522A-3

Alignment Scores:
Pred. No.: 1,7e-12 Length: 2984
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-08-433-522A-3 (1-2984)

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuTleThrGlnGlnGlu 80

Db 923 GAATCTGTTAGTAGCAGTACATTACAAAGCAAAATGGAATTAACAACCTGATTCTTGGTGG 982
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 983 AAATTATGGGAAATAAATTTGAAGGTGCGCAATTCGAGAAAGATTGTCAGTCAATTCGT 1042
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr 116
Db 1043 GATTATTATTAAATAATGCTATGCCAAGCACAAATTAATAAAGCGGTGTCAGCTA 1102
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1103 AATGATGAAAAAACAAGTTAATGTAACCATTTGATGTAATGAAGGTTTACAGTAGTAC 1162
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1163 CTTTCGTAGTCACGCATTTATAGTAAATCTGGGAGGTATGTCGCCGAGCTTGAACCTTTA 1222
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1223 CTTTCAGCATTTAATGATGATACTTTCCGCCGCTAGT----- 1261
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1262 -----GATATTGCAGATGTAGAAAAATGCAATTAAGACAAAACCTTGA--- 1303
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1304 -----GAACGCGGTTACGGTAGCGCAACGGTAAATTCAGTA-----CCT 1342
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1343 GATTTCGTAGTCACAAATAAACATTTAGCGATAACCTTGTTGTTGCTGTCGACGACGT 1402
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1403 TTAACATGTTCCGCCAACCTTCGCTTTGAAGGAAATACCGTTTCTGCTGCTAGTACCTTTACGT 1462
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGln 261
Db 1463 CAGGAAATCGGCCAACAGAAAGAACTTGGTATATAATTCACAAATTTAGTATTGAGTAA 1522
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1523 ATTTCGCTTAGTCGTACAGGTTTCTTCGAA-----ACATCGAAACCAATTTGATCCT 1576
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1577 ATCAATGGTAGTAATGATGAAGTGGATGTCGTATATAAAGTCAAGAACGCTAACACGGGT 1636
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1637 AGTATCAACTTTGGTATTGTTTACGCTACAGAGGTGTATTAGTTATCAAGCAAGTGT 1696
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1697 AAACAAGATAATTTCTTGGAAACAGGGCGGCGAGTAGTATAGTCTGCTACGAAAAATCAT 1756
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1757 TATGTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTACTAAAGATGGTGAAGT 1816
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer-----ThrThrGlnAsn 372
Db 1817 CTTGGTGGAAATGTTTCTTTTGAAGAACTACGATAACTCTAAAAAGTATACATCTCTAAC 1876
QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1877 TATAAGCGTACGACTTACGGAAGTAATGTTACTTTAGGTTTCCTGTAAATGAAATAAC 1936
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402

1262	----	-----GATATCCAGATGTAGAAAATGCAATTAAGACCAAACTTGGA----	1303
QY	186	ValThrArgLysGlyTyProLeuAlaLysLeuGlyAsnThrArgAlaIaValAsnPro	205
DB	1304	-----GAACGCGGTACGGTAGCCCAACGGTAAATTCAGTA-----CCT	1342
QY	206	Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro	221
DB	1343	GATTTTGGATGATGCAATAAACAATACCGATACACCTGTGTGTATGCTGGACGACGT	1402
QY	222	IleAlaPheGlyAspPheGluIleThrGlnArgTyProGluGlnIleValSer	241
DB	1403	TTAACGTGTCGCCAACTTCGCTTGAAGAAATACCGTTTCTGCTGATGACACTTTACGT	1462
QY	242	GlyLeuAlaArgPheGlnProGlyThrProTyArgLeuAspLeuLeuAspPheGln	261
DB	1463	CAGGAATCGCCCAACAAGAACTTGGTATAATTACAAATTAGTTGACTTAGGAAAA	1522
QY	262	GlnAlaLeuGluGlnAsnGlyHisTySerGlyAlaSerValGlnAlaAspPheAspArg	281
DB	1523	ATTTCGGTTAGATGTACAGGTTTCCTCGAA-----ACAGTCGAAACCGCAATGTATCCT	1576
QY	282	LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis	299
DB	1577	ATCAATGGTAGTATGATGAAGTGGATGCTGTATATAAAGTCAAGAACCGTAAACACGGT	1636
QY	300	LysLeuGluThrGlyIleArgLeuAspSerGluTyArgLeuGlyGlyLysIleAlaTy	319
DB	1637	AGTATCAACTTGGTATTGGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGT	1696
QY	320	AspTyTyArgLeuPheAsnLysGlyTyIleGlySerValValTyAspMetAspLys	339
DB	1697	AAACAAGATAATTTCTTGGAAACAGGGCGGCAGTAACTATAGCTGCTACGAAAAATGAT	1756
QY	340	TyrGluThrLeuAlaAlaGlyIleSerGlnProArg-----Asn	353
DB	1757	TATGGTACGAGTGTCAATTTGGTGTATACCGAGCCCTATTTACTAAGATGFGTAAAGT	1816
QY	354	TyrArgGlyAsnTyTyThrThrSerAsnValSerTyArgAsnArgSer---ThrThrGlnAsn	372
DB	1817	CTTGGTGGAAATGTTTCTTGCAAAACACGATAACTCTAAAGTGATACATCTCTTAAC	1876
QY	373	LeuGluLysArgAlaPheSerGlyGlyIle-----TyrTyValArgAspArgAla	389
DB	1877	TATAACGCTACGACTTACGGAAGTAACTGTACTTGTAGTTTCCCTGTAAATGAAATAAC	1936
QY	390	GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu---	402
DB	1937	TCCTATTATGTAGGATTAGGTGCATACCTATATAAAATTTAGTAATTTGCTCTAGAATAT	1996
QY	403	-----GlyArgLysIleProGlySerAsp	410
DB	1997	AACCGTAATTTATATATTCATCAATGAATTAAGGTAAATGCAATTAACCAAAATGAC	2056
QY	411	IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTyLysArgGlnLeuLeu	430
DB	2057	TTTGTAT-----TTTTCTTTTGGTGGAACTATACACGACCTT	2092
QY	431	AsnAsnValLeuHisProGluAsnGlyHisTySerLeuAspGlyLysIleGlyThrLeu	450
DB	2093	ANTAGAGCGTATTTCCCAACTAAAGGG-----GTTAAAGCAAGCTTGTGGTGGACGAGT	2146
QY	451	GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyTyPhePhe	470
DB	2147	ACTATTCCAGGTTCTGATACCAAAATCTACAAACTAAGTGCAGATGTACAGGGTTCTCTAC	2206
QY	471	ThrProGluAsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyThrVal	490
DB	2207	CCATTAGACAGACATCACCTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAA	2266
QY	491	AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer	510
DB	2267	GGTTTTGGAAACAGCGTTTACCGTTCTATCAAACTTATACAGCGGGTGGCATCGGTTCA	2326

RESULT 15

US-08-942-046-3

03 08 242 040 ; Sequence 3, Application US/08942046

; Patent No. 6264954

; PAGE NO: 0204054
; GENERAL INFORMATION:

: APPLICANT: CHONG,

APPLICANT: CHONG, PEI
APPLICANT: THOMAS, WAYNE

APPLICANT: THOMAS, WAYNE
APPLICANT: YANG, YAN PING

APPLICANT: YANG, YAN PING
APPLICANT: LOOSMORE, Sheena

APPLICANT: LOOMORE, Sheena

APPLICANT: STA Dwo Yuan Charles

APPLICANT: SIA, DWO YUAN

APPLICANT: KLEIN, MICHEL
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

	TITLE OF INVENTION:	HANDWRITING
:	NUMBER OF SEQUENCES:	SEQUENCES

NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6TH F

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Can.

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Rel

CURRENT APPLICATION DATA:

APPLICATION

REGISTRATION NUMBER:
FILING DATE:

CLASSIFICATION: 435

CLASSIFICATION: 433
PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: 08/4

APPLICATION NUMBER: 12-SEP
FILING DATE: 12-SEP

;; FILING DATE: 12-SEP-1993
;; CLASSIFICATION: 435

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
NAME: STEWART Michael I

NAME: STEWART, Michael I
REGISTRATION NUMBER: 24 072

REGISTRATION NUMBER: 24,973
REFERENCE DOCUMENT NUMBER: 10

REFERENCE/DOCKET NUMBER:	INTRODUCTION:
1. ;	2. ;
3. ;	4. ;
5. ;	6. ;
7. ;	8. ;
9. ;	10. ;
11. ;	12. ;
13. ;	14. ;
15. ;	16. ;
17. ;	18. ;
19. ;	20. ;
21. ;	22. ;
23. ;	24. ;
25. ;	26. ;
27. ;	28. ;
29. ;	30. ;
31. ;	32. ;
33. ;	34. ;
35. ;	36. ;
37. ;	38. ;
39. ;	40. ;
41. ;	42. ;
43. ;	44. ;
45. ;	46. ;
47. ;	48. ;
49. ;	50. ;
51. ;	52. ;
53. ;	54. ;
55. ;	56. ;
57. ;	58. ;
59. ;	60. ;
61. ;	62. ;
63. ;	64. ;
65. ;	66. ;
67. ;	68. ;
69. ;	70. ;
71. ;	72. ;
73. ;	74. ;
75. ;	76. ;
77. ;	78. ;
79. ;	80. ;
81. ;	82. ;
83. ;	84. ;
85. ;	86. ;
87. ;	88. ;
89. ;	90. ;
91. ;	92. ;
93. ;	94. ;
95. ;	96. ;
97. ;	98. ;
99. ;	100. ;

TELECOMMUNICATION INFORMATION SYSTEMS

Qy	117	-----GluLysAspGlyAla---	TyrThrValHisIleThrProGlyProAsnGlyThrLys	133
Db	1118	AATGATGAAAAACAAAGSTTAATGTGACCATTTGATGTAATGAAGGTTTACAGTATGAC	1177	
Qy	134	-----IleAlaAsnValGly-----	ValAlaIleLeuGlyAspIle	145

[illegible][illegible]

	QY	242	GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuaspLeuLeuaspPheGln	261
	QY	242	GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuaspLeuLeuaspPheGln	261
	Dy	1478	CAGGAATGCCTCAACAAAGAAGAACTGGTATAATTACAATTAGTCAGTTAGAAAA	1537
	QY	262	GlnAlaLeuGluGlnasnGlyHisLysTrSerGlyAlaSerValGlnAlaaspPheaspArg	281
	Dy	1538	ATTGCGTTAGATCGTACAGGTCTTTCGAA-----ACAGTCGAAAACCGAATTCATCCT	1591

Qy	282	LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis	299
	:	:	:
	:	:	:
	:	:	:
Db	1592	ATCAATGGTAGTAAATGATGAAGTGGATGCGTATATAAAGTCAAAAGACGTAACACGGGT	1651
Qy	300	LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleIleTyr	319
	:	:	:
	:	:	:
	:	:	:
Db	1652	AGTATCAACTTTGGTATTGGTATTCACGTCACAGAGTGGTATTAGTTATCAAGCAAGTGT	1711
Qy	320	AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys	339
	:	:	:
	:	:	:
	:	:	:
Db	1712	AAACAAGATATTCTTTGGGAACAGCGGGCGGACGTAAGTATAGCTGGTACGAAAAAATGAT	1771
Qy	340	TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn	353

D6	1172	TAAGGACAGAGATGCAAAATTTGGGTTTATACCGACGCCCTATTTTATATAMGAGATGGTGTAAAGT	1891
Oy	354	TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn	372
D6	1832	CTTGCGGAAGTGGTTTCTTTGGAAACTACGATAACTCTAAAGGTGATACATCCCTCTAAC	1891
Oy	373	LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla	389

DD 1092 TATAAGCGTACGACITACGGAGGTAAATGTTACCTTAGGTTCCCTGTAATAGTAATAAC 1935

D_b 1952 TCCTATTATGTAGGATTAGGTCATACCTATAATAAATTAGTAACCTTTGGCTCTAGAATAAT 2011
||||| :||| |||

QY	403	-----GlyArgLysIleProGlySerasp	410
			:::
Db	2012	AACCGTAATTATATTCAATCAATTAAGGTAATGGCATTAAACAATGAC	2071

Qy 411 IleAspLeuClyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
||| |||
Db 2072 TTTCAT-----TTTCTCTTTGGTTGGAACTATAACAGCCUT 2107

Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2108 AATAGAGGCTATTTCCCAACTAAAGG-----GTAAAGCAAGCTCTGGTGGAGGAGTT 2161
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2162 ACTATTCAGGCTCTGATAACAANAATACTACAACATAAGTCAGATGACAGGGTTTCTAC 2221
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 2222 CCATTACACAGAGATCACCCTCGGTGTATCTGCAANAAGCATCTGCAGGATATGCAAT 2281
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 2282 GGTGTTTGAACAAGCGGTTTACCGTTCTATCAAACTTATACAGCGGTGGCATCGGTTC 2341
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2342 TTACGTGGTTTGTGTTATGAGTAGTATT-----GGACCTAACGCAATTTATGCCGAA 2392
Qy 526 -----GlySerValLeuProGlu 531
Db 2393 TATGTAATGCTAGTGGTACTGTTTAAAGAGATAAGTCTGATGATGTTGGTGGT 2452
Qy 532 ArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2453 AATGCAATCGCTACAGTACAGGCGAGAGTTAATTTGCGCACTCCATTTGTGAGCGATAAG 2512
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2513 AGCCAAATAAGCGTCCGAACCTCTTATTTGTTGATCGCGCAAGTGTGGAATACTAAA 2572
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2573 TGGAAATCAGATAAATAGGATGACAGAGCATGATATTAAMAAATGCGCTGATATGCG 2632
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrrPheSerProLeuAla 584
Db 2633 AAATCAAGCGTATTCGGCGCTCTACAGGTCTCGGATGTCGAATGCCAATCCTCTATTGGG 2692
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2693 CCATTGTTATCTTATGCGCAACCAATTAAMAAATATGAAATGATGATGTCGAA--- 2749
Qy 600 ArgTrrHisLeuSerLeuGlyThrArgPhe 609
Db 2750 CAGTTCCAATTTAGTATGAGGTTCTTTC 2779

RESULT 17

US-09-135-166-55
; Sequence 55, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-135-166-55
Alignment Scores:
Pred. No.: 1.7e-12 Length: 2987
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-09-135-166-55 (1-2987)

Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 938 GAATCTGTTAGTACAGCTACATTTACAAGACAATAATGGAATTTACAACCTGATCTTGGTGG 997
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 998 AAATATGCGGCAATAAATTTGAAGTCCGCAATTCGAGAAAGATTTGCCAGTCAATTCGT 1057
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1058 GATTATTATTTAAATAATATGCTATGCCAAGCACAAATTTACTAAACGGATGTTTCAGCTA 1117
Qy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1118 AATGATGAAAAACAAAAGTTAATGTACCATTTGATGTAATGAAAGTTTACAGTATGAC 1177
Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1178 CTTCGTAGTCACCGCAATATAGGTAATCTGGAGGTATGTCGCGAGCTTGAACCTTTA 1237
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1238 CTTTCAGCATTTACATTTAAATGATACATTTCCCGCGCTAGT----- 1276
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1277 -----GATATTTGCAGATGTAGAAAATGCAATTAAGCAAAACTTTGGA--- 1318
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1319 -----GAACCGCGTTACGGTAGCGCAACCGTAAATTCAGTA-----CCT 1357
Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1358 GATTTTGTATGATGCAAAATAAAACATATAGCGATAACCCCTTTGTTGATGCTGGACGAGT 1417
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGluIleValSer 241
Db 1418 TTAATGTTCCGCCAATCTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGGACTTTACGT 1477

Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGln 261
Db 1478 CAGAAATGCCCAACAAGGAAGTGGTATATAATTCACAAATAGTGGTAGGAA 1537
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1538 ATTGGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAACCGCAATGTATCCT 1591
Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1592 ATCAATGGTAGTAATGATGAAGTGGATGCTATATAAGTCAAGAACGTAACACGGGT 1651
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1652 AGTATCAACTTTGGTATTGGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGT 1711
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLys 339
Db 1712 AAACAAGATATTTCTTGGGAACAGGGCGGCAGTAAGTATAGCTGTACGAAATAATGAT 1771
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1772 TATGGTAGGAGTGCATTTGGGTATACGAGCCCTATTTTACTAAGATGGTGTAGT 1831
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer-----ThrThrGlnAsn 372
Db 1832 CTGGTGGAATGTTTCTTTGGAACACTACGATAACTCTAAAGTGATACATCTCTTAAC 1891
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1892 TATAAGCGTAGCACTACGGAAGTAATGTTACTTTAGTGTTCCTGTAAATGAAATAAC 1951
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1952 TCCTATTATGAGTATTAGGTATACCTATTAATAAATTAAGTCTTGTCTAGATAT 2011
Qy 403 -----GlyArgLysIleProGlySerAsp 410
Db 2012 AACCGTAATTTATATTCAATCAATCAATGAAATTTAAAGTATGCAATTAACAATAACGAC 2071
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2072 TTTGAT-----TTTTCTTTTGGTTGGAACTATAACAGCCTT 2107
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2108 ANTACAGCCTATTCCCAACTAAAGG-----GTTAAAGCAAGCTTTGGTGGACGAGT 2161
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2162 ACTATTCAGGTTCTGTATAACAAATACTACAACTAAGTGCAGATGTACAGGGTTTCTAC 2221
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 2222 CCATTAGACAGATCACCTCTGGTGTGTCTGCAAAAGCATCTGCAGGATATGCAAAAT 2281
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 2282 GGTTTTGGAAACACCGGTTTACCGTTCTATCAAACTATACAGCGGGTGGCATCGTTCA 2341
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2342 TTACGTGGTTTGGTATTAGTATT-----GGACCTAACGCAATTTATGCCGAA 2392
Qy 526 -----GlySerValLeuProGlu 531
Db 2393 TATGGTAATGCTAGTGGTACTTTTAAAGAGATAAGTCTCTGATGTGATTGGTGGT 2452
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2453 AATGCAATCGGTACAGTGGCGAGTTAATTTGGCCAATCCATTCCTGTCAGCGATAAG 2512
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557

Db 2513 AGCAAAATACGGTCCGAACCTCTTATTGTTGATCGCGAAGTGTGGAATACTAAA 2572
Qy 558 -----GlyAspAlaAlaAlaAsnPhelLysArgMet----- 567
Db 2573 TGGAAATCAGATAAAATGATTAGAGAGCGATGTATTAATAAGATTGCCCTGATTATGGC 2632
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAla 584
Db 2633 AAATCAACCGGTATTCCGGCTCTACAGGTGTCGATTCGAATGCCAATCTCCTATTGGG 2692
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2693 CCATTGGTATTCTTATGTCACCAACCAATTAATAAATATGAAATGATGATGTGCAA--- 2749
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2750 CAGTTCCAATTTAGTATTGGAGGTCTTTC 2779
RESULT 18
US-08-942-046-55
: Sequence 55, Application US/08942046
: Patent No. 6264954
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/942,046
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-732 Mts:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2987 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-942-046-55
Alignment Scores:
Pred. No.: 1,7e-12 Length: 2987
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137

DB: 4 Gaps: 24
US-09-857-669-2 (1-609) x US-08-942-046-55 (1-2987)
Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln 80
Db 938 GAATCTGTAGTACAGGTACATTTACAAGACAATATGGAATTTACACACCTGATCTTGGTGG 997
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 998 AAATTTATCGGGAAATAATTTGAAGGTGCGCAATTCGAGAAGATTTTCAGTCAATTCGT 1057
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1058 GATTATTATTTAAATAATAGTCTGCCAAAGCACAAATTAATAAAGCGATGTTCCAGCTA 1117
Qy 117 -----GluLysAspGlyAla-----TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1118 AATGATGAAAAACAAAAGTTTAATGTAACCATTTGATGTAATGAAGGTTTACAGTATGAC 1177
Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1178 CTTCCTAGTCAGCGCATTTATAGTAATCTGGAGGTATGCTCGCGAGCTTGAACCTTTA 1237
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1238 CTTTCAGCATTAACATTAATGATACTTTCCGCCGTAGT----- 1276
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1277 -----GATTTGCGAGATGTAGAAAATGCAATTAAGACAAAACCTTGGAA--- 1318
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1319 -----GAACGGGTTCAGGTAGCCCAACGGTAAATTCAGTA-----CCT 1357
Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1358 GATTTTGTATGATGCAATAAATAACATTTAGCATACCCCTGTTGTTGATGCTGGACGCT 1417
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1418 TTAACCTGTCGCCAACTTCGCTTTGAAGGAATAACCGTTTCTGCTGATAGCAGCTTTACGT 1477
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1478 CAGGAATTCGCCCAACAAGAGAACTTGGTATATATTCACAAATTAGTTGAGTTAGGAAAA 1537
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1538 ATTCGCTTAGATCGTACAGGTTTCTTCGAA-----ACAGTCGAAACCCGAATTCATCCT 1591
Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1592 ATCAATGTAGTATGATGAAGTGGATGCTGTATATAAAGTCAAGAACCGTAACACGGGT 1651
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1652 AGTATCAACTTTGGTATTGGTTACGGTACAGAGAGTGCTATTAGTTATCAAGCAAGTCTT 1711
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1712 AAACAAGATAATTTCTTCGGCAACAGCGCGCGAGTAAGTATAGTGTGTCGAAAAATGAT 1771
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1772 TATGTACAGTGTCAATTTGGTTATACCGAGCCCTATTTTACTAAAGAGTGGTGAAGT 1831
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrTrpGlnAsn 372
Db 1832 CTTGTGGAATGTTTCTTTGAAAAACACGATAACTCTAAAGAGTATACATCCTCTAAC 1891
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389

Db 1892 TATAAGCGCTACGACTTACGGAAGTAATGTTACTTTAGCTTTCCCTGTAAATGAAATAAC 1951
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1952 TCCTATTATGTAGGATTAGGTACATACCTATATAAAATTAGTAATCTTGCCTCTAGAATAT 2011
Qy 403 -----GlyArgLysIleProGlySerAsp 410
Db 2012 AACCGTAATTTATATTCATCAATCAATGAATTTAAAGGTAAATGCAATTTAAACAATGAC 2071
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2072 TTTGAT-----TTTCTTTTGGTTGCACTATAACAGCCCTT 2107
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2108 AATAGAGGCTATTTCCCAACTAAAGGG-----GTTAAAGCAAGTCTTTGGTGGAGAGATT 2161
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2162 ACTATTCAGGTTCTGATAACAATACTACAACACTAAGTCGAGATGTACAGGGTTCTTCTAC 2221
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyThrVal 490
Db 2222 CCATTAGACAGAGATCACTCTGGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAAAT 2281
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyClyAlaSerSer 510
Db 2282 GGTTTTGGAAACAACGCGTTTACCGTTCTATCAAACTTATACAGCGGGTGGCATCGGTGCA 2341
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2342 TTACGTGGTTTGGTTTGTAGTAGTATT-----GGACCTAACGCCAATTTATGCCCAA 2392
Qy 526 -----GlySerValLeuProGlu 531
Db 2393 TATGTAATGCTAGTGGTACTGTTACTTTTAAAGATAAGTCTCTGATGTGATTGGTGGT 2452
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2453 AATGCAATCGGTACAGCTAGCGAGAGGTTAATTTGGCCAACTCCATTTGTGAGCGGTAAG 2512
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2513 AGCCAAATACGGTCCGAACCTCTCTATTCTGTGATGGCGCAAGTGTGTGGAATACTAAA 2572
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2573 TCGAAATCAGATAAAAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2632
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2633 AAATCAACCGGTATTCGGCCCTCTACAGGTGTCGATTTCCCAATGCAATTCCTTCTTATGGG 2692
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysIle 599
Db 2693 CCATTGGTATCTCTTATGCCCAACCAATTTAAAAATATGAAATGATGATGATGATGATGAT 2749
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2750 CAGTTCCCAATTTAGTATTGGAGGTTCTTTTC 2779

RESULT 19

US-09-346-408-5
; Sequence 5, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Anderson, Shawn
; APPLICANT: Falco, Carl
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins


```

; FILE REFERENCE: BB-1167-A
; CURRENT APPLICATION NUMBER: US/09/346,408B
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 60/092,833
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 1697
; TYPE: DNA
; ORGANISM: Glycine max
US-09-346-408-5

Alignment Scores:
Pred. No.: 9,25e-05 Length: 1697
Score: 134,50 Matches: 111
Percent Similarity: 34,46% Conservative: 62
Best Local Similarity: 22,11% Mismatches: 184
Query Match: 4,24% Indels: 146
DB: 25 Gaps: 25

US-09-857-669-2 (1-609) x US-09-346-408-5 (1-1697)

QY 98 AsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer-----Leu 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 AACGCTCTTGTGAACCTTCAGGCGGAGGAGGACGCTTGGGTGGCAGTTTCGAACGCGCTG 292
QY 116 ThrLysAspGlyAlaTyrThrValHisIleThrProGly----- 129
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 ATCGAACCATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 352
QY 130 -----ProArgThrLysIleAlaAsnValGlyValAlaIle 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353 AAGAGGGTGAGGCTCTTTCGTCGCCGAGGATCAAGCTCTCAAGGATGACCTT----- 406
QY 142 LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 -----GAGTGGGTCCCATGTCCTCAGCGAAGGA 433
QY 162 TrpGlnGlnProVal-----GlySerAspPheAspGlnAspSerTrpGlu 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 434 TGGGCCACACCCCTGAAAGGCTTCATGAGAGAACCCGAGTCTCCCAAGCCCTTCATTTC 493
QY 177 AsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAla----- 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 494 AACTCGCTCCGACTCGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGCT 553
QY 195 -----LysLeuGlyAsnThrArgAlaAlaValAsnProAspThr 207
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 ATTGATGATGCGCAGAACGATCGGATCGGGAT-----AACAAAAAGGTT 598
QY 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAla---PheGlyAsp 226
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 GCTCTTTTGTGAT-----TCCAAAGGAGACCCCGTTCGAATTCCTCAATAAT 643
QY 227 PheGluIleThrGlyThrGlnArgTyrPro-----GluGlnIleValSerGlyLeuAla 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 ATTGAGATT-----TATAAGCATCCTTAAGAACAGAAAGATAGCCCAACTTGGGGA 694
QY 245 ArgPheGlnProGlyThrProTyr----- 252
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 695 ACCATTGCGCTGGCTACCTATGTTGAACAAACTATAACCAATGCTGGAATTTGGTTG 754
QY 253 -----AspLeuAspLeuLeu-----LeuAspPheGlnGlnAlaLeuGluGlnAsn 267
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 755 ATTGGGGGTGACCTAGAGGTCAATTGAACCAATTCAGTACATGATGAGCTTGAT----- 808
QY 268 GlyHisTyr-----SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 809 ---CATTTTCTGCTATCTCCGGCACAACTCCGTGTCAGAGTTTCACAAAGCCCAATGCCGAT 865
QY 286 ArgVal-----ProVal-----LysValSerValThr 294
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 866 GCTGTGTTTGCCTCCAGCTCCGGAATCCTGTTCAAAATGCCCATGCTTTGCTAATGACT 925
QY 295 GluValLysArgHisLysLeuGluThrGlyIleArg-----LeuAspSerGluTyrGly 312
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 926 GACACCGGAAGCGCTTCTTGAGTGGCTATGAAGTCTGCTCTCTGCTTCTTCATCCA 985
QY 313 LeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySer 332
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 986 CTTGGAGGC-----TACACCAAGCTGATGATGTCCCA 1018
QY 333 valValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg 352
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1019 CTTGATTGGCAATGAAGCAACACTGAGAGGACTCTTGGAGTGGTGTCTTGATCCAGAG 1078
QY 353 AsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 ACAACTGTGTATCCATATTCCTCATCTCCATGCACTGCTGGACCCAGGAGGTGCAG 1138
QY 373 LeuGluLysArgAlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAsp 392
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1139 -----TGGCATCAAAAGGCTAGG-----ATCAAT 1162
QY 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp-IleAs 412
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1163 GCA-----GGGGCTAACTTCTATATCGTTGGTGGTGGACCCCGCAGCATGCCATCCA 1216
QY 412 pLeuGlyAsnSerHisAlaThrMetLeuThrAla-----SerTrpLysArg 427
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1217 GTTGAGAAAGAGATCTGTATGATGCTGACCATGGAAAGAAAGTATTGAGCATGGCACCG 1276
QY 427 gGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleG 447
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1277 GGAC----- 1280
QY 447 yThrThrLeuGlyThrPheLeuSerSerThrAlaLeuLeuIleArgThrSerAlaArgAlaG 467
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1281 -TAGAGCGCTAAACATCTCTTCCTTCAGGGTTCTGTCATATGACAAGACTCAGGGTAAA 1339
QY 467 yTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaG 487
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1340 ATGGCA-TTCTTTGACCTTCAGGCTCAGGACTTCTGTTCATATCAGCAACAAGAT 1398
QY 487 yTyrThrValAlaArgAsnAlaAspValProSerGlyLeuMetPheArgSerGlyG 507
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1399 GCGCACACTGCCAGGAACAAGAAAGTCTCTCTGATGGATTTATGTCCCTGGTGGATG 1458
QY 507 yAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlyse 527
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1459 GAAGGTGCTGTTGATTACTAT-----GATAGCTTAGTACTCTCAAGC---AACGGCAA 1509
QY 527 rVal 528
Db 1510 AGTG 1513

RESULT 20
US-09-346-408-3
; Sequence 3, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Anderson, Shawn
; APPLICANT: Falco, Carl
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
; FILE REFERENCE: BB-1167-A
; CURRENT APPLICATION NUMBER: US/09/346,408B
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 60/092,833
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 2280

```

```
; TYPE: DNA
; ORGANISM: Momordica charantia
US-09-346-408-3

Alignment Scores:
Pred. No.: 0.00381 Length: 2280
Score: 121.50 Matches: 98
Percent Similarity: 35.70% Conservative: 63
Best Local Similarity: 21.73% Mismatches: 177
Query Match: 3.83% Indels: 113
DB: 4 Gaps: 24

US-09-857-669-2 (1-609) x US-09-346-408-3 (1-2280)
Qy 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
Db 317 CCCAGAAATCGAGCTCTCCAGATCATATCCAGTGGTT---CATGTCTCAGCGAAGC 373
Qy 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProVal----- 166
Db 374 -----TGGCCAGCCCTCTCAGCGGATTC 397
Qy 167 -----GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 398 ATGAGAGAAATCCGAGTCTCTCCAAACGCTTCATTTCAATTCCTCAGGCTCCCGGATGG 457
Qy 185 AlaValThrArgLysGlyTyrProLeuAla-----LysLeuGlyAsnThrArg 200
Db 458 TCTGTGGCGAATATCTCAGTCCCATTTGTTCTCGCCATTGAGGATGCCCAAGACCGC 517
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 518 ATCGCGGACTCCACTTCTGTGCTCTCTTTGAGCGCAAC-----AACAA 562
Qy 221 ProIleAla---PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGln--- 238
Db 563 CCCATCGCCATCTCTCAAGATATTGAGATC-----TACAAGCACCCCTGAAGAGAA 613
Qy 239 ---IleValSerGlyLeuAlaArgPheGlnProGlyThrProTyr----- 252
Db 614 AGGATAGCAAGACCTGGGCGACCACTGCCCGAGGACTCCCTAGTTGATCAACGCTATA 673
Qy 253 -----AspLeuAspLeuLeu-----LeuAspPhe 260
Db 674 ACCAAATGCTGGTAATTGCTGTGATGGGGTGATTTGGAGGTTATAGAACCAATCAAGTAC 733
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 734 CATGATGTCTGGACCGT---TTCCGACAGTCACTCGCAAGCACTACGGGAGGAATTTACC 790
Qy 281 ArgLeuGlnGlyAspArgVal-----ProVal----- 289
Db 791 CGGCGAAATGCGAGATGCGAGTATTTGCAATCCCACTTCGCAATCCATCAACCGGCGAT 850
Qy 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg-----Leu 307
Db 851 GCTTTTACTAATGACTGACACCGCGCTCGGTGCTCGATATGGGCTACAAGAACCCCAT 910
Qy 308 AspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys 327
Db 911 TTGTTGCTTCATCCCTCTGGGTGGC-----TACACCAAG 943
Qy 328 GlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaGly 347
Db 944 GCAGATGATGTTCCTAGTTGGCGAATGAAGCAACATGAGCAAGTCTTCAGATGGT 1003
Qy 348 IleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg 367
Db 1004 GTTCTTGATCCAGACACAACGGTGTCTCCATATTTCCATCTCCAATGCACTATCAGGG 1063
Qy 368 SerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgasp 387
Db 1064 CCAACAGAGGTGCAG-----TGGCATGCAAAAGCG 1093
```

```
Qy 388 ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIlePro 407
Db 1094 AGA-----ATCAATGCA-----GGGCGTAACITTTTACATTTGTTGGCGGTGAC----- 1135
Qy 408 GlySerAspIleAspLeuGlyAsnSerHisAlaThr-----MetLeuThrAla 423
Db 1136 -----CCTGCTGGAAATGGTTCATCTACTAGAAAAGGATCTGTATGATGCT 1183
Qy 424 SerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAsp 443
Db 1184 GACCATGGCAAGAAAGTACTAGCATGGCAGCTGGAGCGGCTAAACATCCTCTCCT 1243
Qy 444 GlyIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 1244 TTCAGAGTT-----GCTGCTTATGATAAACTCAG 1273
Qy 464 AlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg 483
Db 1274 GCGAAAATGGGC---TTTTTCGATCCCTCGAGACCTCAGGATTTCTTTTC---ATATCC 1327
Qy 484 GlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPhe 503
Db 1328 GGCACCAAGATGAGAACACTAGCAAAAGAACAAAGAGAATCCACCAAGAGGTTTATGTGC 1387
Qy 504 ArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGly 523
Db 1388 CCAGGGGATGGAAGGTGTTGTTGAATATTATGAC-----AGTTTGTGCTACCT 1435
Qy 524 ProAsnGlySerValLeuProGluArgAlaLeu 534
Db 1436 GCAAGCAATGACAGACTTCTCTGAACCTGTCTTA 1468
RESULT 21
US-09-346-408-7
; Sequence 7, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Anderson, Shawn
; APPLICANT: Falco, Carl
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
; FILE REFERENCE: BB-1167-A
; CURRENT APPLICATION NUMBER: US/09/346,408B
; EARLIER APPLICATION NUMBER: 60/092,833
; EARLIER FILING DATE: 1999-07-01
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1890
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-346-408-7

Alignment Scores:
Pred. No.: 0.00866 Length: 1890
Score: 117.00 Matches: 110
Percent Similarity: 32.53% Conservative: 52
Best Local Similarity: 22.09% Mismatches: 200
Query Match: 3.69% Indels: 136
DB: 4 Gaps: 26

US-09-857-669-2 (1-609) x US-09-346-408-7 (1-1890)
Qy 104 ArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr 123
Db 112 CGCCGCGCGGCGATGTCGCCCATCCGAGCTCGCTCATCGACCGCGGCGCGCGCTC 171
Qy 124 ValHisIleThr-----ProGly----- 129
Db 172 GTTCGACCTCGTGGCGCCCGGGGAGCCCGCGGCGCTGCGGGCGAGCGGCGGCGGCTC 231
```

```
QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
||||| : : : : : ||| |||
Db 232 CCGGGTGGCTCGCGCGCTGGACGCG----- 261

QY 150 AsnLeuAlaGluTyrThrArgAsnAlaLeuGluAsnTrpGlnProValGlySer--- 168
||||| : : : : : ||| ||| ||| : : : : :
Db 262 -----GAGTGGCGCAGCTGTCGCCGAGGGCTGGCGCTGCCGCTCGCGCTTC 312

QY 169 -----AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
: : : : : ||| ||| ||| ||| |||
Db 313 ATCGGGGACGACGAGTACCTCAGTCCCTCCACCTCACTCCCTCGGCTCCCTCCGCG 372

QY 185 AlaValThrArgLysGlyTyrProLeuAla-----LysLeuGlyAsnThrArg 200
: : : : : ||| ||| ||| ||| |||
Db 373 GGCTCGCCCAACATGTCGCTCCCACTCGCTCGCGCTCGACGACGCCCAAGGACCGC 432

QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
||||| ||| ||| ||| ||| |||
Db 433 GTCGGCGCGCGCGCGACGTCGCG----- 456

QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArg-----TyrPro 236
: : : ||| ||| ||| ||| ||| ||| |||
Db 457 ---CTCGCGCGCGCGCGGAGCTCTCGCGCTCGCGCTCGGAGTCCGAATATACCT 513

QY 237 -----GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyr 252
||||| : : : : : ||| ||| ||| ||| |||
Db 514 CACAATAAAGAAGAGGATTCAACAACATGGGGACAACCTCGCGCTCGCTTACCTTAT 573

QY 253 -----AspLeuAspLeuLeu 257
: : : : : ||| ||| ||| ||| ||| |||
Db 574 GTCGATGAGCGGATACACACGAGTGGAACTGGCTGATTGGTGGTATCTGGAGGTG 633

QY 258 LeuAspPheGlnAlaLeuGluGlnAsn-----GlyHisTyr-----SerGly 272
||||| : : : : : ||| ||| ||| ||| |||
Db 634 -----CAACCCATTAAGTATAACGATGGCTTGACCATTTACAGGCTTTCACCC 681

QY 273 AlaSerValGlnAlaAspPheArgLeuGlnGlyAspArgVal----- 287
: : : : : ||| ||| ||| ||| ||| |||
Db 682 CAGCAACTTAGGACCAATTCGACAGCGTGGGCTGATGCTATTTGCTATTCACGATG 741

QY 288 -----ProVal-----LysValSerValThrGluValLysArgHisLysLeu 301
||||| : : : : : ||| ||| ||| ||| |||
Db 742 AGAAACCCAGTCCACAATGGCATGTCACCTGTTGATGAATGACACTAGAAGGCTCTCT 801

QY 302 GluThrGlyIleArg-----LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
||||| : : : : : ||| ||| ||| ||| |||
Db 802 GAAATGGTTTCAAGATCCCATCTTACTGCTACACCCCTTGGTGGT----- 849

QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
||||| : : : : : ||| ||| ||| ||| |||
Db 850 -----TTTACAAAGCTGATGATGCCGCTGCTGTAGAAATGGAACAA 894

QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
: : : ||| ||| ||| ||| ||| ||| |||
Db 895 CACAGCAAGGCTCTTAAGATGATGAGTCTTGACCCGAGACCACTATCTGTCTATATT 954

QY 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
||||| : : : : : ||| ||| ||| ||| |||
Db 955 CCTCCCCAATGATATTCTGCTGCCAGCAAGTCCAG----- 993

QY 380 GlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
||||| : : : : : ||| ||| ||| ||| |||
Db 994 -----TGGCATCAAGGACCA-----ATTAAGCC-----GGTGTATTTC 1032

QY 400 LeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
||||| : : : : : ||| ||| ||| ||| |||
Db 1033 TACATAGTGGTGGTAT-----CCAGCTGGGATGGCCATCCGACA 1074

QY 420 MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGly 439
||||| ||| ||| ||| ||| ||| |||
Db 1075 GAG-----AAGAGATGCTGTACAAC-----CCAGACCATGGG 1107
```

```
QY 440 ---HisTyrLeuAspGlyLysIleGly---ThrThrLeuGlyThrPheLeuSerSerThr 457
||||| : : : : : ||| ||| ||| ||| |||
Db 1108 AAGAAAGTCTTAAGCATGCCCGGTTTGGAGAACTCAACATATTGCCCTTCAAGGTA 1167

QY 458 AlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeu 477
||||| : : : : : ||| ||| ||| ||| |||
Db 1168 GCAGCATATGATACGCTGGCCAAAGAGATGGCTTTCTTTGAACCTTCACGCGATCAAGAT 1227

QY 478 GlyThrPheIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspVal 497
||||| : : : : : ||| ||| ||| ||| |||
Db 1228 TTTCTGTTCC---ATCTCAGAAACCAAGATGCCACTTTGCCAAAACCTGGAGAACCT 1284

QY 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyr---GluLeu 516
||||| : : : : : ||| ||| ||| ||| |||
Db 1285 CCTGATGGTTTCATGTCGCTGGTGGTGAAGTTCTTTGTTGACTACTACAATAGCTTG 1344

QY 517 AspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeu 534
: : : : : ||| ||| ||| : : : : : ||| |||
Db 1345 CAAACTGAAGGAGTACCGCCCGCGCTGCTACTGTATGAGACAAAGCTGCTG 1398

RESULT 22
5223391-4/c
: Patent No. 5223391
: APPLICANT: COEN, DONALD M.;DIGARD, PAUL E.
: TITLE OF INVENTION: INHIBITORS OF HERPES SIMPLEX VIRUS
: REPLICATION
: NUMBER OF SEQUENCES: 9
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/482,634
: FILING DATE: 21-FEB-1990
: SEQ ID NO:4
: LENGTH: 1654
5223391-4

Alignment Scores:
Pred. No.: 0.0131 Length: 1654
Score: 114.50 Matches: 69
Percent Similarity: 35.60% Conservative: 41
Best Local Similarity: 22.33% Mismatches: 109
Query Match: 3.61% Indels: 90
DB: Gaps: 10

US-09-857-669-2 (1-609) x 5223391-4 (1-1654)

QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
||||| : : : : : ||| ||| ||| ||| |||
Db 1624 GGCACCATCATCCACCCCTTGTATGAGCATCTTACCCCGTAGATGACCGCGATGACTT 1565

QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
||||| : : : : : ||| ||| ||| ||| |||
Db 1564 TTTCTTGGCGAT----- 1553

QY 161 AsnTrpGlnGlnProValGly-SerAspPheAspGlnAspSer-----TrpGluAs 177
||||| : : : : : ||| ||| ||| ||| |||
Db 1552 -----CAGCAGCAGCTTGGTGAACGCTTTTCCGCACTCGAGTTGTATGGGGGAGAA 1499

QY 177 nSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuG 197
||||| : : : : : ||| ||| ||| ||| |||
Db 1498 CAGCGCGCGGAGATGGCTGCCCATC-----TTGTGCCCATGGG 1457

QY 197 yAsnThrArgAlaAlaAsnPro-AspThrAlaThrVal-AspLeuAsnValVal 216
: : : : : ||| ||| ||| ||| ||| |||
Db 1456 CGTCACCGCGCGCGCTGAGCGCGGCGCACACACAAATATGAGTCCGTGTCGCCGTA 1397

QY 217 AspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPro 236
||||| : : : : : ||| ||| ||| ||| |||
Db 1396 GATGATGCGCATGGAATAGGCGCGCGCGCATCTCGCGCGCTCCCGGAAATCGGC 1337

QY 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 256
: : : : : ||| ||| ||| : : : : : ||| |||
Db 1336 CAGGAGCTGTTCGAA-----GGCGCGCCACGCGCGGTGGACGTA 1298

QY 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
```

```
Db 1297 CTCGGCGGTCGCGAGCAGCATCTCGCGCGGATGGTGTGCACCGTCGCGCAACGTGCAG 1238
Qy 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
Db 1237 GCAGCGCAGCAGCTCCGCTGCTGCACTCCGCTGAACCGGTACAC---CGAGTTACA----- 1187
Qy 297 LysArgHisLysLeuGluThrGlyLeuArgLeuAspSerGluTyrGlyLeuGlyGlyLys 316
Db 1186 -----CACGACCTTGATGGCGGCTGCTGCTGTTCAGGAGCAGCGCCTC----- 1142
Qy 317 IleAlaTyrAspTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
Db 1142 ----- 1142
Qy 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 1141 -----CTCGGGGCTGCTCTGGGGAAATCCGCA----- 1115
Qy 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 1114 -----CGCATCTGCTTTCCGATGGCGAGCCAGCTCCCGCAGGAGGATGCTGAGGAG 1064
Qy 377 AlaPheSerGlyGlyIleTyrTyrValArgAspAlaGlyIleAspAlaArgLeuGly 396
Db 1063 GCTCTCTCGCAGCGGCTC-----CTCGGGGCTGCTCTGGGG 1028
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 1027 AATCCGGAGCGGATCTGCTTTCCGATGGCGAGCCAGCTCCCGCAGGAGGATGCTGAGGAG 968
Qy 414 GlyAsnSerHisAlaThrMet 420
Db 967 GCTCTCTCGCAGCTGAGCCTT 947

RESULT 23
US-08-038-682-5
; Sequence 5, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 1 Gaps: 33

US-09-857-669-2 (1-609) x US-08-038-682-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
Db 1529 ATATCGCTAAACACCGGTGGTTT-GTGGAGACGTCGGGGCATGATTATTATTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACGCCAACAGGTGTTGTTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGCATAATGCTATCTATTAAATGCGAGAACACGCA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GGACGC-----AGCAATACTTCAGAAAGACGATGAATACACCGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCGACGACCCCAACCAACGAAACAAACACACATTAACAAACACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATATACTAAAAAAGGTAGC-----TTTCTTAACATCACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TyrGluAsn 177
Db 1810 CGCATCTATGTCATAGCTCCATTAATTTATCCAATGGCAGCTTAACCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GCTCGGAGCGCT-----GCGGCGCT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTAAACAACGATATTACCACCGGTGATGATACCAGAGGTGCAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTACTCAGCGGCTGGTTGATGTTTCATAAAATATCTCACTCGGGGGCGAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GCTAACATAAACATTACAGCTAAACAAGATATCGCTTTTGAGAAGGAACACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGACTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTTAGATTAAATAATGCTCTCTAAACGGCAGCTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGAGTGCAG-----TTTACCACCACTAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
```

Db	2170	AAATAGCGCTTACACAAATAAATTTTCAGAGGACTTTAAATATTTTTCAGGAAAGTCAACATC	2232
Qy	318	-----AlaTyrAsp-----TyrTyr	322
Db	2230	TCAATGGTTTTACCTAAAAATGAAAGTGGATATGATAAATTCAAAGGACGCACCTACTGG	2289
Qy	323	AsnLeuPheAsnLys--GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr	340
Db	2290	AAPTTAAACCTCGAAAGTGGATATGATAAATTCAAAGGACGCCCTCCTACTATTGATCCAGA	2349
Qy	341	GlutThrLeuAlaLaGlyIleSerGlnProArgAsnTyrAlrgYAsnTyrTrpThr	360
Db	2350	GGNAGCGATGTCAGCGCACACTTACCACGCCCTTATAATTTAAACGGT-----	2397
Qy	361	SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly	380
Db	2398	-----ATATCTTCAACAAGACACACTACCTTTTAATGTTGAACGAAATGCA-----	2442
Qy	381	GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly-----	396
Db	2443	-----AGAGTCAACTTGCATCATCAAGGCACCAATAGGGATAAATTAAGTAT	2487
Qy	397	-----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle	411
Db	2488	TCTAGTTTGAATTAACGCATCATTTAATGGAAACATTTTCAGTTTCGGAGGGGGAGTGT	2547
Qy	412	AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn	431
Db	2548	GAITTC-----ACACTCTCGCCTCATCC-----TCT	2574
Qy	432	AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr	448
Db	2575	AACGTCCTCAAAACCCCGGTGTAGTTATAAATCTTAATAATCTTAATGTTCACAGGGTCA	2634
Qy	449	ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaAcqAlaGlyTyr	468
Db	2635	AGTTTAAAGATTAAAACTTCAGGCTCACAAAACACTGGCTCTCAAT-AGAGAAAGATT	2693
Qy	469	PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy	488
Db	2694	AACTTTAAATGCCACCGGAGGCAACATAACACTTTTGCAAGTTGAAGCCACCGATGGAAT	2753
Qy	488	r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr	504
Db	2754	GATTGTGAAGGCATTGTAGCCCAAAAAAACATAACCTTTGAAGGAGGTAAGATG--AG	2810
Qy	504	qSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr	524
Db	2811	GTTTGGCTCCAGGAAGCCGCTAACA-----GAAATCGAAGGCAATGTACTATCAATAA	2864
Qy	524	oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrGl	542
Db	2865	CAACGCTAAGCTC-----ACTCTTATCGGTTTCGGATTTTGACACCATCA	2909
Qy	542	nLeuProPheThr-----ArgThrLeuSerGl	551
Db	2910	AAAAACCTTTAACTATTAAAAAAGATGTCATCATTAATAGCGGCAACCTTACCGCTGGAGG	2969
Qy	551	YAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe	567
Db	2970	CAATATTGTCAATATATACCGCGGAANTCTTACCCTGTCAAAGTAAACGCTAATTTTCAA	3024
Qy	567	tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe	587
Db	3025	-----GCTATCAACAATTTTCAC	3041
Qy	587	rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu-----	605
Db	3042	TTTTTAATGTAGGC--GGCTTGTGTGACAAACAAGGCAATCAAAATATTTCATTGCCAA	3098
Qy	606	----GlyThrArgPhe	609
Db	3099	AGGAGGGCGTTCGCTTT	3114

```

RESULT 24
US-08-302-832-5
; Sequence 5, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; APPLICATION NUMBER:
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 1 Gaps: 33

US-09-857-669-2 (1-609) x US-08-302-832-5 (1-9171)

Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
::: ||||| |||
Db 1529 ATATCGCTAAACCGGTGGTTT-GTGGAGACGTCGGGGCATGATTTATTCATCAAGAC 1587
::: |||||

Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGlu 86
::: ||| ::::| ::::|
Db 1588 AATGCA-----ATTGTTGACGCAAGAGTGGTTGTTAGC----- 1623
::: |||||

Qy 87 GlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
::: |||||
Db 1624 -----CCGGAATATGTTATCTATTAAATGCAAAACACGA 1656
::: ||| ::::|

Qy 107 GlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
::: ||| ::::|

```

Db 1657 GGACGC-----AGCAATATCTCAGAAGACGATGAATACACGGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCACGCCCAACAGCAAGAAACAAAGAACACATTAACAACACACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATACTAAAAAGGTACC-----TTTGTACATCATCTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspClnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCATAGTCCATTAATTTATPCCAATGCGACTTAACTCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTCCGAGCGGT-----GGCGGCGTT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTAAACACGATATTACCACCGCTGATACACAGAGGTGCAAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTTACTCAGCGCGCTGGGTTGATGTTTCATAAAATATCTCAGTCGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAACATTTACAGCTAAACAAGATATCCGCTTTGAGAAGGAAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGCTCAAGGACTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAGGTTTATAGATTAAATGATCTCTCTAAAGCGCACTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTACCACTAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AATACGCTATCACAAATAAATTTGAAGGACTTTAAATATTTTCCAGGAAAGTCAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGTTTTTACCTAAAAATGAAAGTGGATATGATATAAATTCAAAGGACCGCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTTAACCTCGAAGTGGATATGATAAATTCAAAGGACCGCTCACTATTGATCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTCAGGCGACACTTACCAGCGCTTATAATTTAAACGGT----- 2397
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTTCAACAAAGACACTACCTTTAATGTTGAACGAAATGCA----- 2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGNGTCAACTTTGACATCAAGGCACCAATAGGATAAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTGAATTACGCATCATTTAATGTAACATTTTCAAGTTCGGGAGGGGGAGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTC-----ACACTTCTCGCCTCATCC-----TCT 2574

Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AACGTCCAAACCCCGGTGTAGTTATAAATCTCTAATTAATCTTTAATGTTTCAACAGGTCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 2635 AGTTTAAGATTAAACCTTCAGGCTCAACAAAAAACTGGCTTCTCAAT-AGAGAAAGATT 2693
Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTyr 488
Db 2694 AACTTTAAATGCCACCGAGGCAACATAACACTTTTCAAGTTTGAAGGCACCGATGAAT 2753
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGGTAAAGGCATTGTAGCCAAAAAACATACCTTTGAAGGAGGTAAAGATG---AG 2810
Qy 504 gSerGlyGlyAlaSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGGAAGCCGTAAACA-----GAAATCGAAGGCAATGTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrGl 542
Db 2865 CAAGCGCTAACGCTC-----ACTCTTATCGGTTTCGGATTTTGACAAACCATCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerGl 551
Db 2910 AAAACCTTTAACTATTAAAAAAGATGTCATCATTAATAAGCGCAACCTTACCGCTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgme 567
Db 2970 CAATATTGTCATATAGCGGAATCTTACCGTTGAAAGTAACGCTAATTTCAAAA----- 3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAAATTTTCAC 3041
Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTATGTAGGC---GCCTGTGTTTGACAAACAAAGGCAANTTCAANTTTCCATTGCCAA 3098
Qy 606 ---GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114
RESULT 25
US-08-530-198-5
; Sequence 5, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

Db 3025 -----GCTATCACAATTTCCAC 3041
Qy 597 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAATGAGGC---GCCTGTTTGGACAACAAGGCNAATTCAAATATTTCCATTGGCAA 3098
Qy 606 ----GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114

RESULT 26
US-08-469-880-5
; Sequence No. 5876733
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-469-880-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66

Db 1529 ATATCGCTAAACCGGTGGT---GTGAGACGTCGGGGCATGATTTATTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACGCCAAAGAGTGGTGTGTTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGGTAATGTATCTATTAATGTCAGAACACGCA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GGACGC-----AGCAATATCTCAGAAGACGATGAATACACCGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCAGCACCCCAACAAAGAAACAAAGAACACATTAACAAACACAACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATATAAAAAAGGTACC-----TTTGTTAACATCAGCTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCAATAGCTCCATTAATTTATCCAATGCGAGCTTAACCTCTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GCTCGGACCGGT-----GGCGCGGT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GACATTAAACACGATATTACCACCGGTGATGATACCAGAGGTGCAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTACTCAGCGCGGTGGTGTGTTGATGTTCAATAAATATCTCATCGGGCGGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAAACATTACAGCTAAACAGATATCGCCCTTTGAGAAGGAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuLeuLeuAsp 259
Db 2059 ATTACAGTCAAGGAGCTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTATAGATTTAATAATGCTCTCTAAACGGCAGCTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTCCACCCTAAAGAAGCAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AATACGCTATCACAAATAAATTTGAAGGGACTTTAATAATTTTCAGGGAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTGTTCCTAAATAATGAAAGTGGATATGATAAATTCAAAGGACGCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTrp 340
Db 2290 AATTTAACTCGAAAGTGGATATGATAAATTCAAAGGACGCCCTCAGTACTGCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTGCAGGCACACATTTACCAGCGCTTATAATTTAAACGGT----- 2397
Qy 361 SerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPheSerGly 380

Db 2398 -----ATATCATTTCAACAAGACACTACCTTTAATGTTGAACGAATGCA----- 2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCACTTTGACATCAAGCGCAATAGGATAAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTGAATTACGCATCATTTAATGGAAACATTTCACTTCGGGAGGGGAGGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTC-----ACACTTCTCGCCCTCATCC-----TCT 2574
Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTrpLeuAspGlyLysIleGlyThr 448
Db 2575 ACCTGCAAAACCCCGGTGAGTATATAATTTCTAAATACATTTTAATGTTTCAACAGGATCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyThr 468
Db 2635 AGTTTAAGATTTAAACTTCAGGCTCAACAAAACTCGCTTCTCAAT-AGAGAAGATTT 2693
Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
Db 2694 AACTTTAAATGCCACCGAGGCAACATAACACTTTTGCAAGTTGAAGCCACCGATGAAT 2753
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGGTAAAGGCATTTAGCCAAAAAACAATAACCTTTGAAGGAGGTAAAGATG----AG 2810
Qy 504 gSerGlyGlyAlaSerSerValArgGlyTyrgluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGGAAACCGGTARCA-----GAATCGAAGGCATGTTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrgl 542
Db 2865 CAACGCTAAGCTC-----ACTCTTATCGTTCGGATTTTGACAAACCATCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerGl 551
Db 2910 AAACCTTTTAACTATTAAAAAGATGTCATCATTAATAGCGGCAACCTTACCGGTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
Db 2970 CAATATTGCTCAATATACCGGAATCTTACCCTTGAAGTAACGCTAATTTTCAA----- 3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaPropheSe 587
Db 3025 -----GCTATCACAAATTTTAC 3041
Qy 587 rPheAspIleAlaTyrglyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAATGTAGGC---GGCTTTGTTGACAAACAAAGCAATTCAAATATTTTCCATTGCCAA 3098
Qy 606 ---GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114

RESULT 27

US-08-728-470-5

Sequence 5, Application US/08728470

Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 9171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-728-470-5

Alignment Scores:

Pred. No.:	0-23	Length:	9171
Score:	114.00	Matches:	138
Percent Similarity:	32.82%	Conservative:	74
Best Local Similarity:	21.36%	Mismatches:	235
Query Match:	3.59%	Indels:	201
DB:	2	Gaps:	33

US-09-857-669-2 (1-609) x US-08-728-470-5 (1-9171)

Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
Db 1529 ATATCGCTAAACCGGTGGTGT-TGGAGAGCGTCGGGCATGATTTTATTCATCAAAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACGCCAAAGAGTGGTGTGTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGATTAATGTATCTATTATATGCGAAGACAGCA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyThr----- 123
Db 1657 GGAGCG-----AGCAATACTTCAGAGAGCATGAATACACGGGATCCGG 1701
Qy 124 ---ValHisIleThrProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCAGCACCCCAACAAACAAAGAACACATTAACACAAACACAACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyThrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATACTAAAAAAGGTACC-----TTTGTAAACACTACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCATAGCTCCATTAATTTATCCAAATGCGACGCTTAACCTCTTTGGAGTGAG 1869

```
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTGGAGCGGT-----GGCGGCGTT-----1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTTAACAGCATATTTACCACCGGTGATGATACCAGAGGTGCAAAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----Alaphe 224
Db 1939 TTAACAATTTACTCAGCGGCTGGTGTGATGTTCTATAAAAATATCTCACTCGGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAAATACAGCTAAACAGATATTCGCCCTTTGAGAAAGGAAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGACTATTACCTCAGGCAAT-----2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTAGATTTAATGTTCTCTAAACGGCAGCTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTCAACCACTAAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIle----- 317
Db 2170 AAATACGCTATCACAAATAAATTTGAAGGGACTTTAAATATTTTCAGGGAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTTTTACCTAAATAAGTGGATATGATATAATTCAAAGACGCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys-----GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTAACTCTGAAAGTGATATGATAATAATTCAAAGGAGCGCCCTCACTATTGACTCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCATAGTCAGGCGACACTTACCAGCGCTTATAATTTAAACGGT-----2397
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTCACAAAGACACTACCTTTAATGTTGACCAAAATGCA-----2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly-----396
Db 2443 -----AGAGTCAACTTTGACATCAAGGCACCAATAGGATAAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTGAATACGCGATCATTTAATGGAACAATTTCACTTCGGAGGGGGAGGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTTC-----ACACTTCTCGCCTCATCC-----TCT 2574
Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AACGTCCAACACCCGGTGTAGTTATAAATTTCTAAATACTTTAATGTTTCAACAGGTCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 2635 AGTTTAAGATTTAAACTTCAGGCTCAACAAAAAAGTCTGCTCAAT-AGAGAAAGATTT 2693
Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
Db 2694 AACTTTAATGCCCGGAGGCAACATAACACTTTTTCAGCTTCAAGGACCGCGATGNAAT 2753
```

```
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGGTAAAGGCATTTGTAGCCAAAAAACATACCTTTTGAAGGAGGTAAAGATG---AG 2810
Qy 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGGTCCAGGAAGCCGTAACA-----GAAATCGAAGGCAATGTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrG1 542
Db 2865 CAACGCTAACGTC-----ACTCTTATCGGTTCCGATTTTGCACACCATCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerG1 551
Db 2910 AAAACCTTTAACTATTAAAAAAGATGTCATCATTAATAGCGCAACCTTTACCGCTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
Db 2970 CAATATTGTCAATATAGCGGAAATCTTACCGTTTGAAGTAACGCTAATTTTCAAA-----3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----CCTATCACAAATTTTCAC 3041
Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysIleArgTrpHisIleSerLeu-----605
Db 3042 TTTTAATGTAGGC---GCCTTGTGTTGACAAACAAAGCAATTTCAAAATATTTCCATTGGCAA 3098
Qy 606 ----GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114
RESULT 28
US-08-617-697-5
: Sequence 5, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
```

LENGTH: 9171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-5

Alignment Scores:

Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Watch: 3.59% Indels: 201
DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-617-697-5 (1-9171)

```
Qy 47 VallysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
   ::::::::::::::::::::
Db 1529 ATATCGCTTAAACCGGTGGT---GTGAGACGTCGGGGCATGATTTATTCATCAAGAC 1587
   ::::::::::::::::::::
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluGluValLeuAspLysGlu 86
   ::::::::::::::::::::
Db 1588 AATGCA-----ATTGTCAGCGCCAAAGAGTGGTGTGTTAGAC----- 1623
   ::::::::::::::::::::
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
   ::::::::::::::::::::
Db 1624 -----CCGATTAATGTATCTATTAATGCGAGAACAGCA 1656
   ::::::::::::::::::::
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
   ::::::::::::::::::::
Db 1657 GGACGC-----AGCAATACTTCAGAAGACGATGAATACACCGGATCCGGG 1701
   ::::::::::::::::::::
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
   ::::::::::::::::::::
Db 1702 AATAGTCCGACGCCCAACAAAGAAAGAACACATTAACAAACACAACTCTT 1761
   ::::::::::::::::::::
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
   ::::::::::::::::::::
Db 1762 GAGACTATCTAATAAAGGTACC-----TTTGTTAACATCACTGCTAATCAA 1809
   ::::::::::::::::::::
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
   ::::::::::::::::::::
Db 1810 CGCATCTATGTCATAGCTCCATTAATTTATCCAAATGCGCAGCTTAACCTTTGGAGTGAG 1869
   ::::::::::::::::::::
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
   ::::::::::::::::::::
Db 1870 GTCGAGCGGT-----GGCGCGGT----- 1890
   ::::::::::::::::::::
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
   ::::::::::::::::::::
Db 1891 -----GAGATTAAACAGATATTACCACCGGTGATGATACAGAGGTGCAAC 1938
   ::::::::::::::::::::
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
   ::::::::::::::::::::
Db 1939 TTAACAATTTACTACGCGCGTGGTGTGATGTTTCAATAAATATCTACTCGGGCGCAA 1998
   ::::::::::::::::::::
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
   ::::::::::::::::::::
Db 1999 GTTAACATAAACATTACAGCTAAACAGATATCGCCTTTTGAGAAAGAGCAACCAAGTC 2058
   ::::::::::::::::::::
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
   ::::::::::::::::::::
Db 2059 ATTACAGGTCAGGCACTATTACCTCAGGCAAT----- 2091
   ::::::::::::::::::::
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
   ::::::::::::::::::::
Db 2092 -----CAAAAAGGTTTAGATTAAATAATGTCCTCTAAACCGGCACTGGC 2136
   ::::::::::::::::::::
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
   ::::::::::::::::::::
Db 2137 AGCGACTGCAA-----TTCAACCACTAAAGAACCAAT 2169
   ::::::::::::::::::::
```

```
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIle----- 317
   ::::::::::::::::::::
Db 2170 AATACGCTATCACAAATAAATTTGAAGGGACTTTAAATATTTTCAGGAAAGTGAACATC 2229
   ::::::::::::::::::::
Qy 318 -----AlaTyrAsp-----TyrTyr 322
   ::::::::::::::::::::
Db 2230 TCAATGCTTTTACCTAAAAATGAAAGTGGATATGATATAAATTTCAAAGGACGCACTACTGG 2289
   ::::::::::::::::::::
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTyrAspMetAspLysTyr 340
   ::::::::::::::::::::
Db 2290 AATTTAACTCGAAAGTGGATATGATATAAATTTCAAAGGACGCGCTCCTACTATTGATCCAGA 2349
   ::::::::::::::::::::
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
   ::::::::::::::::::::
Db 2350 GGAAGCGATAGTCGAGCGACACATTACCAGCCTTATAATTTTAAACGCT----- 2397
   ::::::::::::::::::::
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysAlaPheSerGly 380
   ::::::::::::::::::::
Db 2398 -----ATATCATTTCAACAAGACACTACCTTTAATGTTGAACGAATGCA----- 2442
   ::::::::::::::::::::
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
   ::::::::::::::::::::
Db 2443 -----AGAGTCACCTTTGCACATCAAGGACCAACCAATAGGATAAATAAGTAT 2487
   ::::::::::::::::::::
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
   ::::::::::::::::::::
Db 2488 TCTAGTTTGAATTACGCATCATTTAATGGAACATTTTCAGTTTCGGGAGGGGAGTGT 2547
   ::::::::::::::::::::
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
   ::::::::::::::::::::
Db 2548 GATTTC-----ACACTTCTCGCCTCATCC-----TCT 2574
   ::::::::::::::::::::
Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
   ::::::::::::::::::::
Db 2575 AACGTGCAACCCCGGTGTAGTTATAAATTTCTAAATCTTTAATGTTTCAACAGGGTCA 2634
   ::::::::::::::::::::
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
   ::::::::::::::::::::
Db 2635 AGTTTAAGATTAAACTTTCAGGCTCAACAAAACACTGCTCTCAAT-AGAGAAAGATTT 2693
   ::::::::::::::::::::
Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
   ::::::::::::::::::::
Db 2694 AACTTTAAATGCCACCGGAGGCAACATACACTTTTGAAGTTGAAGGACCGCATGGAAT 2753
   ::::::::::::::::::::
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
   ::::::::::::::::::::
Db 2754 GATTGGTAAAGCGCATTTGAGCCAAAACAAAACATACCTTTGAAGGAGGTAAAGATG---AG 2810
   ::::::::::::::::::::
Qy 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
   ::::::::::::::::::::
Db 2811 GTTTGGCTCCAGGAAAGCGGTAAACA-----GAAATCGAAGCAATGTTATATCAATAA 2864
   ::::::::::::::::::::
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrG1 542
   ::::::::::::::::::::
Db 2865 CAACGCTAACGTC-----ACTCTTATCGTTCGGATTTTGAACAACCATCA 2909
   ::::::::::::::::::::
Qy 542 nLeuProPheThr-----ArgThrLeuSerG1 551
   ::::::::::::::::::::
Db 2910 AAAACCTTTAACTATTAAAAAAGATGTATCATTAATAGCGGCAACCTTACCGTGGAGG 2969
   ::::::::::::::::::::
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
   ::::::::::::::::::::
Db 2970 CAATATTCTCAATATAGCGGAAATCTTACCGTTGAAGTAACGCTAATTTCAAA----- 3024
   ::::::::::::::::::::
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaPropheSe 587
   ::::::::::::::::::::
Db 3025 -----GCTATCACAAAATTTTCAC 3041
   ::::::::::::::::::::
Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
   ::::::::::::::::::::
Db 3042 TTTTAATGTAGC---GGCTTGTGACAAACAAAGGCAATTCAAATATTTTCCATTGCCAA 3098
   ::::::::::::::::::::
Qy 606 ---GlyThrArgPhe 609
```

```
Db 3099 AGGAGGGGCTCGCTTT 3114
|||||
RESULT 29
US-08-719-641-5
; Sequence 5, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-719-641-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
Db: 4 Gaps: 33

US-09-857-669-2 (1-609) x US-08-719-641-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
Db 1529 ATATCGCTAAACCGGTGGTTTTT-GTGAGAGCGTCGGGGCATGATTTATTTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACGCCAAAGAGTGGTTGTTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106

Db 1624 -----|||||
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GGAGCC-----AGCAATACTTTCAGAAAGACGATGATCAATACCGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCAGCACCCCAACAAAGAAAGACAAACATTAACAAACACAACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 CAGAGTATATCTAAAAAAGGTACC-----TTTGTTAACATCCTCTAATCA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCATAGCTCCATTAATTTTATCCATGGCAGCTTAACCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTGGAGCGGT-----GGCGGCGTT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTACACAGCATATTACCACCGGTGATGATACCAGAGGTGCAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTACTCAGCGCGGTGGTTGATGTTCAATAAAATATCTCACTCGGCGCGCA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAACATTACAGCTAAACAGATATCGCCTTTGAGAAAGGAAAGCAACAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGAGCTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTATAGATTTTAAATATGCTCTCTAACGGCAGCTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTCAACACTAAAAAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AAATACGCTATCACAAATAAAATTTCAAGGGACTTTTAAATATTTCAAGGAAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTTTACCTAAATAATGAAGTGGATATGATAAATCAAGAGCGGACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTAACTCGAAGTGGATATGATAAATCAAGAGGAGCGGCTCACTATTGACTCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGGATAGTGCAGGCACACTTACCAGCGCTTATATTTAAACGGT----- 2397
Qy 361 SerAsnValSerTyrAspArgSerThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTCACAAAGACACTACCTTTATGTTGAACGAATATGCA----- 2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCAACTTTGACATCAAGGCACCAATAGGATAAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
||| |||
```



```
Db 439 GTTTATTATGAATAAGTAAGATATTATCA----- 468
Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
Db 469 -----AGGATATTAAAGTAAATTT-----AATCAACCATATCAG 504
Qy 254 -----LeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyrSer 271
Db 505 AAATTTTATGATGATTAAATACCAATTAATAATGATCTGATTCAGATGGCAA----- 558
Qy 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValproVallysVal 291
Db 559 -----GATCTTTTA----- 567
Qy 292 SerValThrGluVallysArgHisLysLysLeuGluThrGlyIleArgLeuAspSerGluTyr 311
Db 568 TTTACTAATCAGCTTAAGAACATCCACAGACATTTTCTGTAGAATTTCTTGAACAAAT 627
Qy 312 GlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGly 331
Db 628 -----AGCAATGAGGTACAAGAAGTATTGCGAAAGCTTTT----- 663
Qy 332 SerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro 351
Db 664 ---GCATATTATATCGAGCCACAGCATCGTGATGTTTACAGCTTTATGCACCGGAAGCT 720
Qy 352 ArgAsnTyrArgGlyAsnTyrTriphThrSerAsnValSerTyrAsnArgSerThrThrGln 371
Db 721 TTTAATTACATGGATAAATTTAACCAACAAGAAATAATCTAACCGCTGCGGCCACGCC 780
Qy 372 AsnLeuGluLys-----ArgAlaPheSer 379
Db 781 GACGTGTGAGCCTGACCTGCCGCTGCCGCGGTGAATGCGCGGCGCGGACAGC 840
Qy 380 GlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeu----- 395
Db 841 GGC-----GACGCCCTGCTGGAGCGCAACTAT 867
Qy 396 -----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 868 CCCACTGCGCGGAGTTCTCGCGGCGCGCGCGGCGGACGTC----- 906
Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrp-----LysArgGlnLeuLeuAsn 431
Db 907 -----AGCTTCAGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Qy 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
Db 961 CACGCCCAACTGGAGGCGCGGCTATGTGTGCTGCGGTACCCACGCGGCGGCTCTCCGAA 1020
Qy 452 ThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr 471
Db 1021 -----GCGGCCCAAGAGTCTCTCGCGGCGGCTGCGCGG----- 1056
Qy 472 ProGluAsnLysLysLeuGlyThr-----PheIleIleArgGlyGln----- 485
Db 1057 ---CGCAGCCAGGACTCGACGCGATCTGCGCGGCTTCTATATCGCGCGGATCCGCGC 1113
Qy 486 AlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer 505
Db 1114 CTGGCCTACCGCTACGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1170
Qy 506 GlyGly-----AlaSerSerValArgGlyTyrGluLeuAspSer 518
Db 1171 GGTGCTCTGCTCGGGTCTATGTGCGCGCTCGAGCCTGCGCGGCTTCTACCGCAGCAGC 1230
Qy 519 IleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer 538
Db 1231 CTGACCTGTGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1281
Qy 539 LeuGluTyrGlnLeuPro 544
Db 1282 -----CATCGCTGCGC 1293
```

```
RESULT 31
US-08-082-849B-9
: Sequence 9, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: TITLE OF INVENTION: Related Methods
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082,849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus anthracis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1524
: OTHER INFORMATION: /product= "LF(1-254)--TR--PE(362-613)"
US-08-082-849B-9
```

```
Alignment Scores:
Pred. No.: 0.0189 Length: 1524
Score: 112.50 Matches: 112
Percent Similarity: 34.80% Conservative: 78
Best Local Similarity: 20.51% Mismatches: 189
Query Match: 3.55% Indels: 167
DB: 1 Gaps: 28
```

US-09-857-669-2 (1-609) x US-08-082-849B-9 (1-1524)

```
Qy 38 LysAsnLysSerProAspThrGluSerValLysLeuLysProLysPheProValArgIle 57
Db 40 AAAGAGAAAAATAAAGATGAGCAAT-----AAGAGAAAGATGAAGAACGAAT 87
Qy 58 AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGln 77
Db 88 AAAACACAGGAAGAGCAATTTAAAGGAATCATG---AAACACATTTGTAATAAATAGAGTA 144
```

```
Qy 78 GlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAsp 97
   :: : ||||| :: : ||||| ||| ||||| |||
Db 145 AAAGGGAGGAAGTGTAAAAAGAGCGACGACGAGAAAGCTACTTGGAGAAAGTACCATCT 204
   :: : ||||| :: : ||||| ||| ||||| |||
Qy 98 AsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGlu 117
   :: : ||||| :: : ||||| ||| ||||| |||
Db 205 GATGTTTAGAGATGTATAAAGCAATTTGGA-----GGAAGATATATATTGTG--- 252
   :: : ||||| :: : ||||| ||| ||||| |||
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLys 133
   ||||| ||||| ||||| ||||| |||||
Db 253 ---GATGGTGATATTACAAACATATATCTTTAGAACATTATCTGAAGATAAGAAAANA 309
   ||||| ||||| ||||| ||||| |||||
Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu 153
   ||| :: : ||||| ||||| ||||| |||||
Db 310 ATAAAGAC-----ATTATGGG-----AAAGATGCTTTATTATACATGAA 348
   ||||| ||||| ||||| ||||| |||||
Qy 154 TyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGluAsp 173
   :: : ||||| ||||| ||||| ||||| |||||
Db 349 CATTTATGATATGCAAGAAGAGATATGACCCCTACTTGTATCCCAATCTTCGGAAGAT 408
   ||||| ||||| ||||| ||||| |||||
Qy 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu 193
   ||||| ||||| ||||| ||||| |||||
Db 409 TATGTAGAAATACTGAAAGGCA-----432
   ||||| ||||| ||||| ||||| |||||
Qy 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
   ||||| ||||| ||||| ||||| |||||
Db 433 -----CTGAAC 438
   ||||| ||||| ||||| ||||| |||||
Qy 214 ValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGln 233
   ||||| ||||| ||||| ||||| |||||
Db 439 GTTTATTATGAAATAGTAGATATATATCA-----468
   ||||| ||||| ||||| ||||| |||||
Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
   ||||| ||||| ||||| ||||| |||||
Db 469 -----AGGGATATTTAAGTAAATTT-----AATCAACCATATACAG 504
   ||||| ||||| ||||| ||||| |||||
Qy 254 -----LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSer 271
   ||||| ||||| ||||| ||||| |||||
Db 505 AAATTTTATGATGATTAAATACCATTAATAATGCATCTGATTCAGATGGACAA----- 558
   ||||| ||||| ||||| ||||| |||||
Qy 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysVal 291
   ||||| ||||| ||||| ||||| |||||
Db 559 -----GATCTTTT-----567
   ||||| ||||| ||||| ||||| |||||
Qy 292 SerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyr 311
   :: : ||||| ||||| ||||| ||||| |||||
Db 568 TTTACTAATCAGCTTAAGAACATCCACACACTTTTCTGTAGAAATCTTCGGAACAAAT 627
   ||||| ||||| ||||| ||||| |||||
Qy 312 GlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGly 331
   :: : ||||| ||||| ||||| ||||| |||||
Db 628 -----AGCAATGAGGTACAAGAAGTATTTCGCAAGCTTTT----- 663
   ||||| ||||| ||||| ||||| |||||
Qy 332 SerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro 351
   :: : ||||| ||||| ||||| ||||| |||||
Db 664 ---GCATATTATTCGAGCCACAGCATCGTGATGTTTACAGCTTTATGCAACCGGAAGCT 720
   ||||| ||||| ||||| ||||| |||||
Qy 352 ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln 371
   ||||| ||||| ||||| ||||| |||||
Db 721 TTTAATTACATGGATAAATTTTAACGACAAAGAAATAAATCTAACCGGTGGCGGCAACGCC 780
   ||||| ||||| ||||| ||||| |||||
Qy 372 AsnLeuGluLys-----ArgAlaPheSer 379
   :: : ||||| ||||| ||||| ||||| |||||
Db 781 GACGTGGTGAGCCTGACCTGCCGCTGCCGCGGTGAATGCGCGGGCGCGGACAGC 840
   ||||| ||||| ||||| ||||| |||||
Qy 380 GlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeu----- 395
   ||||| ||||| ||||| ||||| |||||
Db 841 GGC-----GACGCGCTGCTGGAGCGCAACTAT 867
   ||||| ||||| ||||| ||||| |||||
Qy 396 -----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
   ||||| ||||| ||||| ||||| |||||
Db 868 CCCACTGGCGGAGTTCTCTGGCGACGGCGGCGACGTC-----906
   ||||| ||||| ||||| ||||| |||||
```

```
Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrp-----LysArgGlnLeuLeuAsn 431
   ||||| ||||| ||||| ||||| |||||
Db 907 -----AGCTTCAGCACCCGCGCACGAGAACTGGAGCTGGAGCGCTGCTCCAGCGC 960
   :: : ||||| :: : ||||| ||| ||||| |||
Qy 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
   :: : ||||| ||||| ||||| ||||| |||||
Db 961 CACCGCCAACTGGAGGAGCGCGGTATGTGTCGCTACACGCGACCTTCTCTCGAA 1020
   :: : ||||| ||||| ||||| ||||| |||||
Qy 452 ThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr 471
   :: : ||||| ||||| ||||| ||||| |||||
Db 1021 -----GCGGGCAAGCATCGTCTTCGCGGGGTGCGCGC----- 1056
   :: : ||||| ||||| ||||| ||||| |||||
Qy 472 ProGluAsnLysLysLeuGlyThr-----PheIleIleArgGlyGln----- 485
   :: : ||||| ||||| ||||| ||||| |||||
Db 1057 ---CGCAGCCACCTCGACCGCATCTGCGCGGTCTATATCTCGCGCGATCCGCGC 1113
   :: : ||||| ||||| ||||| ||||| |||||
Qy 486 AlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer 505
   ||||| ||||| ||||| ||||| |||||
Db 1114 CTGGCCTACGGCTACGCCAGCAGCAGACCCACGCGCGC-----CGATCCGCAAC 1170
   ||||| ||||| ||||| ||||| |||||
Qy 506 GlyGly-----AlaSerSerValArgGlyTyrGluLeuAspSer 518
   ||||| ||||| ||||| ||||| |||||
Db 1171 GGTCCCTGCTGCGGGTCTATGTCCGCGCTCGAGCTGCGCGGCTTCTACCGCACCGC 1230
   ||||| ||||| ||||| ||||| |||||
Qy 519 IleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer 538
   :: : ||||| ||||| ||||| ||||| |||||
Db 1231 CTGACCTGGCGCGCGGAGCGCGGCGGAGTCTCGAACGG-----CTGATCGGC--- 1281
   ||||| ||||| ||||| ||||| |||||
Qy 539 LeuGluTyrGlnLeuPro 544
   :: : ||||| ||||| ||||| ||||| |||||
Db 1282 -----CATCGCTGCGC 1293
   ||||| ||||| ||||| ||||| |||||
RESULT 32
PCT-US94-01624-9
: Sequence 9, Application PC/TUS9401624
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: TITLE OF INVENTION: RELATED METHODS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
: STREET: Steuart Street Tower, 20th Floor, One Market
: STREET: Plaza
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/01624
: FILING DATE: June 25, 1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-115
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
```

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /product=
PCT-US94-01624-9 "LF(1-254)--TR--PB(362-613)"

Alignment Scores:
Pred. No.: 0.0189 Length: 1524
Score: 112.50 Matches: 112
Percent Similarity: 34.80% Conservative: 78
Best Local Similarity: 20.51% Mismatches: 189
Query Match: 3.53% Indels: 167
DB: 5 Gaps: 28

US-09-857-669-2 (1-609) x PCT-US94-01624-9 (1-1524)

Qy 38 LysAsnLysSerProAspThrGluSerValLysLeuLysProLysPheProValArgIle 57
Db 40 AAAGAGAAAATAAGATGAGAAAT-----AAGAGAAAAGATGAAGAAAGAAAT 87
Qy 58 AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGln 77
Db 88 AAACACAGGAGGACATTTAAAGGAATCATG---AAACACATTTGTAANAATAGAACTA 144
Qy 78 GlnGlnGluGluValLeuAspGluGlnThrGlyPheLeuAlaGluGluAlaProAsp 97
Db 145 AAAGGGAGGAGGAGCTGTTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 204
Qy 98 AsnValLysThrMetLeuSerSerLysGlyTyrPheSerSerLysValSerLeuThrGlu 117
Db 205 GATGTTTTAGAGATGATAAGCAATTCGA-----GGAAGATATATATATGTC--- 252
Qy 118 LysAspGlyAlaTyrValHisIleThr-----ProGlyProArgThrLys 133
Db 253 ---GATGCTGATATACAAACATATATCTTTAGAGCATTTATCTGAAGATAAGAAAAA 309
Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu 153
Db 310 ATAAAGAC-----ATTATGGG-----AAAGATGCTTTATATACATGAA 348
Qy 154 TyrTrpArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAsp 173
Db 349 CATATGATATGCAAAAGAGGATATGACCGCTACTTGTAAATCCCAATCTTCGGAAGAT 408
Qy 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu 193
Db 409 TATGTAGAAAATCTAGAAAGGCA-----432
Qy 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
Db 433 -----CTGAAC 438
Qy 214 ValValValAspSerGlyArgProfileAlaPheGlyAspPheGluIleThrGlyThrGln 233
Db 439 GTTTATTATGAATAGGTAGATATATCA-----468
Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
Db 469 -----ACCAACCATATACAG 504
Qy 254 -----LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSer 271
Db 505 AAATTTTATGATGATTAATACCATTAATAAATGCATCTGATTCAGATGGACAA----- 558
Qy 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysVal 291
Db 559 -----GATCTTTTA-----567

Qy 292 SerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyr 311
Db 568 TTTACTAATCAGCTTAAGGAACATCCACACAGACTTTCTGTAGAATTTCTTGAACAAAT 627
Qy 312 GlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGly 331
Db 628 -----AGCAATGAGCTACAAGAAGATTTTGCAGAAAGCTTTT----- 663
Qy 332 SerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro 351
Db 664 ---GCATATTATATCGAGCCACAGCATCGTGTGTTTACAGACTTTATGACCCGGAAGCT 720
Qy 352 ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgThrThrGln 371
Db 721 TTTAATTACATCGGATAAATTTAACGACACAGAAATAAATCTAACCGCTGCGGCCACGCC 780
Qy 372 AsnLeuGluLys-----ArgAlaPheSer 379
Db 781 GACGTGTGAGCTGACCTGCCGCGGTGCGCGGTGAATGCGCGGGCGCGCGGACAGC 840
Qy 380 GlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu----- 395
Db 841 GGC-----GACGCCCTGCTCGAGCGCAACTAT 867
Qy 396 -----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 868 CCCACTGCGCGGAGTCTCTCGGCGAGCGCGGCGACGTC----- 906
Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrp-----LysArgGlnLeuLeuAsn 431
Db 907 -----AGCTTCAGCACCGCGGCGACGAGAACTGGACGGTGGAGCGCTGCTCCAGGCG 960
Qy 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
Db 961 CACGCCCAACTGGAGGCGCGGTATGTTGTCGCGCTACCGAGGCACTTCCTCGAA 1020
Qy 452 ThrPheLeuSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr 471
Db 1021 -----GCGGCGCAAGCATCGCTCTCGCGGGGTGCGCGG----- 1056
Qy 472 ProGluAsnLysLysLeuGlyThr-----PheIleIleArgGlyGln----- 485
Db 1057 ---CGCACCCAGGACCTCGACGCGATCTGGCGCGGCTTTCTATATCGCGCGGATCCGCGC 1113
Qy 486 AlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer 505
Db 1114 CTGGCTACGGCTACGCCCCAGGACCGAACCAGGACCGCGCGG-----CGGATCGGCAC 1170
Qy 506 GlyGly-----AlaSerSerValArgGlyTyrGluLeuAspSer 518
Db 1171 GGTGCCCTGCTCGGGTCTATGTGCGCGCTCGAGCTCGCGGCTTCTACCGCACCAAGC 1230
Qy 519 IleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer 538
Db 1231 CTGACCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1281
Qy 539 LeuGluTyrGlnLeuPro 544
Db 1282 -----CATCCGCTGCCG 1293

RESULT 33
US-08-363-124A-1
; Sequence 1, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated

STREET: 180 Varick Street
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10014
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 FILING DATE: US/08/363,124A
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/124,254
 FILING DATE: 20-SEP-1993
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/973,336
 FILING DATE: 05-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/572,187
 FILING DATE: 23-AUG-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gallagher, Thomas C.
 REGISTRATION NUMBER: 37,066
 REFERENCE/DOCKET NUMBER: SPA-1-PDC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 645-1405
 TELEFAX: (212) 645-2054
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3286 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:
 ORGANISM: Neisseria gonorrhoeae
 STRAIN: FA19
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 406..3150
 US-08-363-124A-1

Alignment Scores:
 Pred. No.: 0.0732 Length: 3286
 Score: 112.00 Matches: 150
 Percent Similarity: 34.09% Conservative: 89
 Best Local Similarity: 21.40% Mismatches: 281
 Query Match: 3.53% Indels: 181
 DB: 2 Gaps: 36

US-09-857-669-2 (1-609) x US-08-363-124A-1 (1-3286)

QY 85 -----LysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnVally 100
 Db 633 CCGGACCTGACGGTTACGACCCCGCATCGCGTCTCGAACAGGCGCGCGCAAG 692
 QY 100 sThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspG1 120
 Db 693 CTCGGGCTACTCGATACGGGTATGGACAAAAACCGGCTCTCCTTGACG---GTGACGG 749
 QY 120 y-----AlaTyrThrValHisIleThrProGlyProArgThrLysIleAl 135
 Db 750 CTTGGGCCAAATACAGTCTTACACCGCAGCGGCATTTGGGC---GGGACGAGGACGGC 806
 QY 135 aAsnValGlyValAlaIle-----LeuGlyAspIleLe 146
 Db 807 GGGCAGCAGCGCGCAATCAATGAAATCGACTATGACAGCTCAAGCTCTGCGAATCAG 866
 QY 146 uSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn-----TrpG1 163
 Db 867 CAAAGGCTCAAACTCGGTCGGAACAAGCAGCGGCATTTGGCGGCTCGGTTCGCAATTCA 926
 QY 163 nGlnProValGlySerAspPhe-----AspGlnAspSerTrpGlu---AsnSerLysTh 180
 Db 927 AACCAAAACCGCGACGATGTTATCGGGGAAGCAGCAGTGGGCGATTCAGATGAATAAC 986
 QY 180 rSerValLeuGly-----AlaValThrArgLysGlyTyrProLeuAlaLysLeuG1 197
 Db 987 CGCTATTTCGGGCAAAAACCGGGGCTTACCCAATCATCGCGCTGGCGGGCGCATCGG 1046
 QY 197 yAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValas 217
 Db 1047 CGGTGCGGAGGCT-----TTGCTGATCCG 1070
 QY 217 pSerGlyArgProIleAlaPheGlyAspPheGluIleThrClyThrGlnArgTyrProG1 237
 Db 1071 CACCGCGCGCACGCG-----GGGAAATCCGCGCCACGAGCCCGCGACGCGCGT 1124
 QY 237 uGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLe 257
 Db 1125 TCAGAGCTTCACAGCGCTGGCGCGTGTGATGACGGCAGCAAGTACGCTATTTCATCGT 1184
 QY 257 uLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAl 277
 Db 1185 T-----GAAGAAGATGCAAAAACGGGGGTAC----- 1212
 QY 277 aAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLly 297
 Db 1213 -----GAAAAGTGTAAAGCGAAT-----CCGAAAAAAGATGTTGTCGCGCAAGACAA 1259
 QY 297 sArgHisLysLeuGluThrGlyIleArgLeuAspSerGlyTyrGlyGlyLysI1 317
 Db 1260 ACGTCAACGGTTTCCACCCGAGACTACACGGGCCCCAACCGCTTCTCGCGCATCCGCT 1319
 QY 317 eAlaTyrAspTyrTyrAsn---LeuPheAsnLysGly----- 328
 Db 1320 TTCATAGAAAGCGCGTGGCTGTTCCGCGCGGTTTTCGTTTGTAAACAAACAGCGCA 1379
 QY 329 -TyrIleGlySerValVal-----TrpAspMetAspLysTyrGluThrThrLe 344
 Db 1380 CTACATCGCGCGCATACTCGAACGACGCAACAACTTTCGACACGCGCGCATGACGGT 1439
 QY 344 uAlaAlaGlyIleSer-----GlnProArgAsnTyrAr 355
 Db 1440 TCCGCGCATTTCTGACCAAGCGGTTTTTGATGCAAAATCAAAACAGGCGGTTCTTTCGCG 1499
 QY 355 gGlyAsn---TyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuG1 374
 Db 1500 CGGCAACGGCAAAATACGCGGGCAACCAAAATACGCGGAGCTGTTTACCAGCGGC---GA 1556
 QY 374 uLysArgAla-----PheSerGlyGlyIleTyrValArgAspArgAl 389
 Db 1557 AAACAATGCGCGGTGGGCGCGGAATACGTACGCGGCTGTTTAC-----GACGAGAC 1610

QY 6 ThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyrAlaProAlaAlaAsp 25
 Db 411 ACAGCAACATTTGTCGGATTAATAATTTATGCTGTCTTAATGACTGCGCTGCCGCG 470
 QY 26 Leu-SerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrG1 45
 Db 471 TTATGCGAATAATGTGCAAGCGGCAAGCA-----CAGAAAAACAGTTGGATACCAT 524
 QY 45 uSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLy 65
 Db 525 ACAGTAAAGCCCAAAACAGAAACCCCGCGC-----GATAACGAGTAAC 572
 QY 65 sAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluGluValLeuAsp-- 84
 Db 573 CGGTTTGGGCAAAATTTGGTCAAAACCGCGCACACCTCAGCAAGGAACAGGTACTCGACAT 632

```
Qy 389 aGlyleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlyse 409
Db 1611 GCACACAAAGCGCTACGTTGGGAATATGTC-----TATACCAATGC 1655
Qy 409 rAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLe 429
Db 1656 CGATAAAGACACTTGGCGGATTATGCCGCTC-----TCTTACGACCGGCAG-- 1704
Qy 429 uLeuAsnAsnValLeuHisProGluAsnGlyHisTyLeuAspGlyLysIleGlyThrTh 449
Db 1705 -----GGCATCGTTTGGCAACACCATTTTCAGCAGAC 1736
Qy 449 rLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTrpH 469
Db 1737 GCACTGTTCCGCGCAGCGTTTCGGCAAAATATTCGCGCGAGTCCGCAAGCGGTTTC 1796
Qy 469 ePheThrProGluAsnLysLeuGlyThrPheIleIleArgGly-----484
Db 1797 CTATTTACAAATCCGACGCG-----GTGATTTACGGGAAAGCCATAAGCT 1841
Qy 485 ---GlnAlaGlyThrValAlaAArgAspAsnAlaAspValProSerGlyLeuMetPh 503
Db 1842 CTTCAGCGCGCATTCAAAAATCTTCGATACCGCCCAAAATCCGCCAACCTTGAGCGT 1901
Qy 503 eArgSerGly-----GlyAla-----se 509
Db 1902 GAATCTCGTTACGACCGCTTCGGCTCTAATCTCCGCCCATCAGGATTATTATTATCAAG 1961
Qy 509 rSerValArgGlyTyrrGluLeuAsp-----517
Db 1962 TGCCAACCGCGCTATTCTGTTGAAACGCCCTCTCAAAACAACGCGCAAAAAACAGCC 2021
Qy 518 -----SerIleGlyLeuAlaGlyProAsnGlyse 527
Db 2022 CAACGGCAGAGAAAGAAATCCCTATTGGGTCAGCATAGGC-----AGGGGAAA 2069
Qy 527 rValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrrGlnLeuProPheThr 547
Db 2070 TGCTGTTACAGGCAAAATCTCGCTTTTGGCAACAATACITATACGCACTGCCGCGCG 2129
Qy 547 gThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMe 567
Db 2130 CAGCATCAACGGCAAAAGCTATTAC-----GC 2156
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 2157 GCGGTCGGGCAACAAATCTCGCTTTGGGC---AGTGCGGCGATCTCGCGCGGCTTGGC 2213
Qy 587 rPheAspIleAlaTyrrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyTh 607
Db 2214 CTACGACTACCGCAGCAGCATTCGGACGAC-----GGCAGCGCTTCCACCGCGCAC 2264
Qy 607 r 607
Db 2265 G 2265
```

RESULT 34

```
US-08-565-386-10
; Sequence 10, Application US/08565386
; Patent No. 5741697
; GENERAL INFORMATION:
; APPLICANT: Bavoil, Patrik M.
; APPLICANT: Hsiao, Ru-Ching
; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; STREET: Clinton Square, P.O. Box 1051
; CITY: Roches
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/565,386
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 176/60040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
US-08-565-386-10
Alignment Scores:
Pred. No.: 0.0314 Length: 1791
Score: 111.50 Matches: 107
Percent Similarity: 35.60% Conservative: 55
Best Local Similarity: 23.52% Mismatches: 214
Query Match: 3.51% Indels: 79
Gaps: 22
```

US-09-857-669-2 (1-609) x US-08-565-386-10 (1-1791)

```
Qy 14 PhePhePheProHisAlaTyrrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGly 33
Db 352 TATTTTGTCTCTCAAGTAACCTTCCTCGATGGAGGTTACGCGAGAAATCTATTATTATGAC 411
Qy 34 PheAlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhe 53
Db 412 TATTTTGGTCTACCTACTAAAGTTGCAAAATTATCGGCAT---CAAGTTTTCCTCCATTAAGA 468
Qy 54 ProValArgIleAspThrGlnAspSerGluLysAspMetValGluGluHisLeuPro 73
Db 469 GCATATAATTTGATATTAAATAGATATATGATGATGATGATGATGATGATGATGATGATGAT 528
Qy 74 LeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu 93
Db 529 GTTTGGACAGGAGATGCTGATCTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTT 585
Qy 94 Glu-----AlaProAspAsnValLysThrMetLeuArgSerLys-----GlyTyr 108
Db 586 GAGCATGATGACAGTCTCCTTATGATATATAAGTTAAGTTAAGTTAAGTTAAGTTAAGTTAAGTT 645
Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrrThrValHisIleThrPro 128
Db 646 TTTACTTCTGCTCTTCTGCTGTTTGCAGAAAGGTCCTTCTGTTGGAATAGGATTATTACAGGT 705
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIle-----LeuGlyAspIle 145
Db 706 GGAGATTACAGGAGCTCTCCAGTTTCATGTTTACGATTTAGATCTTATTGGATGATTTCT 765
Qy 146 LeuSerAspGly---AsnLeuAlaGluTyrrArgAsnAlaLeuGluAsnTrpGlnGln 164
Db 766 TCTGATGATCATGTTTGTAGTTTGTGTTGTTCTTATGATTAACGCTTCCACAGAAATGG----- 819
Qy 165 ProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 820 -----TTTACTGCAGAT-----GGTCTGTTGACTTCTTCTGGAATGGGT 855
Qy 185 AlaVal-----ThrArgLysGlyTyrrProLeu-----AlaLys 195
Db 856 AGTGTCTCTGTTGGTACAACTGGTAATTTCTCTATGATTAATGTTGTTGATCCATCTTAT 915
Qy 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
```

Db	916	TTTGGTACACACTGTTGCCCAAACTGGTAGTCATCTTCTTCT	-----	957
QY	216	ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyr	235	
Db	958	-----TCTACTCCGCCTTTGTTTAAAGGTGATTTTCTCTGTT	-----TAT	996
QY	236	ProGluGlnIleValSerClyLeuAlaAargPheGlnProGlyThrProTyrAspLeuAsp	255	
Db	997	GTTGAT-----TTACGGCTTCATCTTTCAGTTACGATTAATTCGTTCTCGT	1041	
QY	256	LeuLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyrSerGly	-----	272
Db	1042	AATGCCATTAATCTTCCACACAGTGGTTTGAGAAGAGTGCTCTTATGGAAGTAGATATGTT	1101	
QY	273	AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp	-----ArgValProVal	289
Db	1102	GAATCTGTTCAAGGTCAATTTTGGCTTCATCTTGGTGATTAATCGTCTACGACCAATC	1161	
QY	290	-----LysValSerValThrGluValLysArgHisLysLeuGlu	302	
Db	1162	TATTTAGGTGGATCTAAGTCTTATGTTTCTGTTAATCTCGTAGACAGAATTC	-----	1215
QY	303	ThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr	-----	319
Db	1216	-----TCTACAGATTCAGTTTCTCCTCAAGGAATCTTCTGCTTATGCATTAATCT	1266	
QY	320	---AspTyrTyrAsnLeuPheAsnLysGlyTyrIle	-----GlySerValValTirPasp	336
Db	1267	ACAGATACTAAACATTTGTTTACGAAGCTTTTGTGTAGCATGGTTTGTATAGGTCCT	1326	
QY	337	Metasp-----LysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro	351	
Db	1327	CTTTAGCTACAGCGGATTTAACTTATCAGCAAGGTTTATAGAGCTCAGTGGTCAAGATT	1386	
QY	352	ArgAsnTyrArgGlyAsnTyrThrPheSerAsnValSerTyrAsnArgSerThrThrGln	371	
Db	1387	AGTCGGTAT---GATTATTTTGGCTACTTTTGTCTATTTGGGAGAGACGCCGTGTTAT	1443	
QY	372	AsnLeuGluLysArgAlaPheSerGlyGlyIleThrTyrValArgAspArgAlaGlyIle	391	
Db	1444	AATAAAGAGATTTATTGCCAATCAGATACCTGTTATGGATCCTAGTGGTCTGCGGTTAAT	1503	
QY	392	AspAlaArgLeuGly---AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp	410	
Db	1504	GATTCGCCTTTGGTTATCAAGAGCGTTATGCTGAGTATCGTTATAAGCCTTCGAAGTT	1563	
QY	411	IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTyr	425	
Db	1564	ACTGGATTATTTAGATCTAACGCTACAGGTACTCTAGATTCTTGG	1608	
RESULT 35				
US-09-090-793-13				
; Sequence 13, Application US/09090793				
; Patent No. 6140486				
; GENERAL INFORMATION:				
; APPLICANT: Calgene, LLC				
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression				
; OF INVENTION: of polypeptide-like synthesis genes in plants				
; FILE REFERENCE: CGNE.131.01US				
; CURRENT APPLICATION NUMBER: US/09/090,793				
; CURRENT FILING DATE: 1998-06-04				
; EARLIER APPLICATION NUMBER: 60/048,650				
; EARLIER FILING DATE: 1997-06-04				
; NUMBER OF SEQ ID NOS: 66				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 13				
; LENGTH: 19227				
; TYPE: DNA				
; ORGANISM: Vibrio marinus				
US-09-090-793-13				
Alignment Scores:				
Pred. No.:	1.41	Length:	19227	

Score:	111.50	Matches:	77
Percent Similarity:	35.2%	Conservative:	55
Best Local Similarity:	20.5%	Mismatches:	137
Query Match:	3.51%	Indels:	105
DB:	3	Gaps:	15
US-09-857-669-2 (1-609) x US-09-090-793-13 (1-19227)			
Qy	76	ThrGlnGlnGlnGluValLeuLeuAspGlyGlnThrGlyPheLeuAlaGluGluAla	95
Db	9162	ACCGCATTCGAGCAAGCTACGATTAAGTTCGAGAGCGTTGCTTTGTTAGAGTGTCAACA	9221
Qy	96	---ProAspAsnValLysThrMetLeuArgSerLysGly---	110
Db	9222	GTCGCTGATTCGGCAATCGCATGTGCTGAAGACCAAGCTTTAATGTCTGCTTCATCATCAT	9281
Qy	111	SerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrProGlyPro	130
Db	9282	ACGCAAACTTTGCACTACTGCATTAGCAGTGCCTGCTGACTGTGAGGCGGGTGT	9341
Qy	131	ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn	150
Db	9342	TTTTCACAGTCCGACGGTTATTGAATGTGTAATGTG	9380
Qy	151	LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPhe	170
Db	9381	TTACATCAACGTTATATTCGGCGATTAAAGATTGGCAACAACCG-----AGTGACAAT	9434
Qy	171	AspClnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGly	190
Db	9435	CAAAATGCACGGTGGCGGAATTCA---CCATTCTATATGCCTGTAGATGCTCGACCTGG	9491
Qy	191	TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal	210
Db	9492	TTCCCATCTGCTGATCGCTCTGCACACATTCGCGCTTATAGTTGTGTGACTGCTGCACAGC	9551
Qy	211	AspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr	230
Db	9552	TATTGTCATATTCTTTTACAA-----	9572
Qy	231	GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaAArgPheGlnProGlyThr	250
Db	9573	-----GAAACGCTTTACAGAACCTGTT---TTGAAAGAAACAGCTC	9611
Qy	251	ProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln-----	266
Db	9612	TTGCAAGTAAATGACTTAACTGAAACAGCTTTCAGACTCTTGACAAACAAATCCAGTA	9671
Qy	267	-----AsnGlyHisTyrSerGly-----	272
Db	9672	GCTGATCTGCGCACTAATGTTACTTTGCATCGACGAGTTAGCATTAATCATAGTACAA	9731
Qy	273	-----AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValPro	288
Db	9732	GGTAATGACGAACACAAATACGCTGTGAATTTAGAACTATTACAGGCGACTTAAGTACT	9791
Qy	289	ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp	308
Db	9792	ACTGGCAATAGTACTATCAGTATATAA-----	9818
Qy	309	SerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyr-----AsnLeuPheAsn	326
Db	9819	-----CAGATCGCAGCAGACTGTTATGCTCCCGTAATGATACTAAC	9857
Qy	327	LysGlyTyr-----IleGlySerValValTrpAspMetAspLysTyrGluThr	342
Db	9858	AAAGCCATATACGCAGCTGCTTATTTCGCGAGACTGCTGAAGAGTTAAGCAAAATAAC	9917
Qy	343	ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr-----	354
Db	9918	TTGCGGTTGCTGGTATCGCTAGCTGGTGTAAATGAAGATGCTAAGAAATGGAACCCCG	9977
Qy	355	ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg-----SerThrThrGlnAsn	372

Alignment Scores:
Pred. No.:

```

Db 9978 AAGGGCAGTATTATTTACCGCGCAGCCCTGCAATAAACAGGCTGCTAACAGCACAGCAAT 10037
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIlePrpTyrValArgAspArgAlaGlyIleAsp 392
Db 10038 GGTGTCACCTTCATGTCACCGGTT----- 10061
Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 10062 -----ATTGCTGCTACATATGTTGGTTTAGGGCGTGATCTA 10097

RESULT 36
US-09-090-793-12
; Sequence 12, Application US/09090793
; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; FILE OF INVENTION: of polyketide-like synthesis genes in plants
; FILE REFERENCE: CGNE.131.0105
; CURRENT APPLICATION NUMBER: US/09/090.793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 40138
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-12

Alignment Scores:
Pred. No.: 4.58 Length: 40138
Score: 111.50 Matches: 77
Percent Similarity: 35.29% Conservative: 55
Best Local Similarity: 20.59% Mismatches: 137
Query Match: 3.51% Indels: 105
DB: 3 Gaps: 15

US-09-857-669-2 (1-609) x US-09-090-793-12 (1-40138)
Qy 76 ThrGlnGlnGluGluValLeuAspGluGlnThrGlyPheLeuAlaGluGluAla 95
Db 26145 ACCGCATTGCAGCAAGTAGCATAACTGCAGCAGCGTGGTTGTTAGAGTGTCTACGA 26204
Qy 96 ---ProAspAsnValLysThrMetLeuArgSerLysGly-----TyrPheSer 110
Db 26205 GTCGCTGATTGGCGCAATCGCATCTCTGAAAGCCCAAGGTTTAATGTCTGCTTATCATCAT 26264
Qy 111 SerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHsIleThrProGlyPro 130
Db 26265 ACGCAACTTTCATCTACTGCAATTAAAGCAGTCCCGTAGTGTGCTGAGGCGGGTGT 26324
Qy 131 ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn 150
Db 26325 TTTTCACAGGTCGCGAGGTTTATTCAAATGTGTAATTGCT----- 26363
Qy 151 LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPhe 170
Db 26364 TTATCAACGCTTATATCCGGCCATTAAGATTGGAACAACCG-----AGTGACAAT 26417
Qy 171 AspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGly 190
Db 26418 CAATGTTCACCGTCGCGGAATTCA---CCATTCTATATGCTGTAGATGCTCGACCTTGG 26474
Qy 191 TyrProLeuLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
Db 26475 TTCCACATGCTGATGCTCTGCGACACATGCGCCTTATAGTTGTGTGCTGCTGACAGC 26534
Qy 211 AspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
Db 26535 TATTGTATATCTCTTTTACAA----- 26555
```

```

Qy 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThr 250
Db 26556 -----GAAACGCTCTTACAAACAACCTTGTT---TTCAAAGAAACAACAGTC 26594
Qy 251 ProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln----- 266
Db 26595 TTGCAAGATATAGCTTAACCTGAAAGCAAGCTTCAGACTCTTGAACAAACAACATCCAGTA 26654
Qy 267 -----AsnGlyHisTyrSerGly----- 272
Db 26655 GCTGATCTGCGCACTAATGTTACTTTGTCATCGAGCGAGTTAGCATTAAATCATAGTACAA 26714
Qy 273 -----AlaSerValGlnAlaAspPheAspArgLeuGlnGlnArgValPro 288
Db 26715 GGTAAATGACGACGACCAATTAAGCTGTGAATTAGAACTATTACAGGCGCAGCTTAAGTACT 26774
Qy 289 ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp 308
Db 26775 ACTGGCATAGTACTATCATGATTAA----- 26801
Qy 309 SerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyr-----AsnLeuPheAsn 326
Db 26802 -----CAGATCGCAGCAGACTGTTATGCCCTAATGATGACTAACT 26840
Qy 327 LysGlyTyr-----IleGlySerValValTrpAspMetAspLysTyrGluThr 342
Db 26841 AAGCCTATAGCGCAGTGTCTATTGCGGAGACTGCTGAGAGTTAAGCAAGAAATACCC 26900
Qy 343 ThrLeuAlaGlyIleSerGlnProArgAsnTyr----- 354
Db 26901 TTGGCGTTTGTGCTATCGCTAGCGTGTATTAAATGAAGATGCTAAAGAAACCCCG 26960
Qy 355 ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg-----SerThrThrGlnAsn 372
Db 26961 AAGGCGAGTATTATTTACCGCGCAGCTGCAAAATAACAGGCTGTCTAACAGCACACAGTA 27020
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAsp 392
Db 27021 GGTGTCACCTTCATCTACCCAGGT----- 27044

RESULT 37
US-08-164-292B-15
; Sequence 15, Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
```

Qy	206	AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly	225
Db	2665	GAA---GGGCAACTGCAGCTAAACATTACTCGCGGTGAGGCGCTCAACTTGTTCGAAACAAC	2721
Qy	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
Db	2722	AGCTCGCGGTG-----GAGCTGGCGCTCGGCGCTC---CAT	2754
Qy	246	PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu-----	256
Db	2755	TTTCCCGCTGGCCAAACCAAGTAAGCCTTTATCCCGAGATGGAATAGACATCCCGAGAT	2814
Qy	257	-----LeuLeuAspPheGlnGlnAlaLeu	264
Db	2815	AATAGGTGACTGTCGCGCTGGCGCAGCGCTGAGATGCTCAACCAACCAACTTGCCTG	2874
Qy	265	GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly	284
Db	2875	GCT-----TCCGAGACGGTTTGAAGTCCACACGCGACCCCTC-----	2913
Qy	285	AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu	301
Db	2914	---CGGTAAAGCTCTCCACCGCCTGACATTTGAAATGGCGCGTACGAGCAAACTA	2970
Qy	302	GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly	315
Db	2971	GGACGAGACTTGGCACACAGCACTGCTGTCGTCGCTGGTTTCGCACAGGTCGAGGACTT	3030
Qy	316	LysIleAlaTyrAspTyrTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal	333
Db	3031	AGAGTTGCAACGCCCAAGTCCAGATCTTTCAGCGGAAGAGCAAGCGCATCGGCACCT---	3087
Qy	334	ValTrpAspMetAspLysTyrGluThrThrLeuAlaGlyIleSerGlnProArgAsn	353
Db	3088	---GATAGCAGCCTCACTCTCAACATCCGGGCGCGCCCTCAAA	3126
Qy	354	TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu	373
Db	3127	TTTTCTGGACCGCCTTGACTGCTAGTTTGCAGGACGAGTGGTCCGATTACTTACAACAGC	3186
Qy	374	GluLysArgAlaPhe-----SerGlyGlyIleTrpTyrValArgAspArgAla	389
Db	3187	AACAATGGCACTTTCGCTCTCTATAGCCCGCGGAATGTGGTAGACACAAACAGACTT	3246
Qy	390	GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAla	401
Db	3247	CAGGTAACCCAGCGCGCTGTTAGTCTTCCAAGGAAACAACCTTGTCCCAACCTTGGC	3306
Qy	402	GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer	416
Db	3307	GATCCGCTGCTATTTCGACACGCAAAATAGTCTCAGTCTCGTCCGCGGCTGACCCAA	3366
Qy	417	HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro	436
Db	3367	GCTTCCAACGCCCTGACTTTAAGT-----TTAGGAACCGGCTTGAATTC	3411
Qy	437	GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer	456
Db	3412	TCCAAT---CAAGCCGTTGCTATAAAGCGCGCGGCGCTTA---CGCTTTTGGTCTCTCC	3465
Qy	457	ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys	476
Db	3466	TCACAGCTTTAGAGAGCGCTCACAGTCGGAAATGGCTTAACG-----	3510
Qy	477	LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn---	494
Db	3511	CTTACCAGTACTGTGATCCGCCCAACCTAGGCGGCGCTAGAGGTCTCAGAGACAATAAA	3570
Qy	495	-----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArg	512
Db	3571	ATCATTTTAAGTGGCGCGAATCTTCGTTTTTGAACGGAGCGCGCTAACCGCC-----	3624

Alignment Scores:					
Pred. No.:	0.19	Length:	5100		
Score:	111.00	Matches:	146		
Percent Similarity:	34.04%	Conservative:	78		
Best Local Similarity:	22.19%	Mismatches:	260		
Query Match:	3.50%	Indels:	174		
DB:	1	Gaps:	36		

US-09-857-669-2 (1-609) x US-08-164-292B-15 (1-5100)

Qy	3	IleLysProThrAlaLeuLeu-----	LeuProAlaLeuPhePheProHis	18
Db	2131	ATTAAACCGCGGGCTTCTGGAGACAACGACCTCGCGTGCTCTCCACC	CGGTG	2190
Qy	19	AlaThrAlaProAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys	38	
Db	2191	GCCTCCGATGAGCGCGCAAGTCACGCTCAACATGTCTGACGGCTATATACTAAGGAC	2250	
Qy	39	AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe-	-----	53
Db	2251	AACAAGCTA-----	-GCTGTCAAAGTAGGTCCCGGCTGTCCTCGACTCCCAAT	2298
Qy	54	---ProValArgIleAspThrGlnAsp-----	SerGluIleLysAspMetVal	68
Db	2299	AATGCTCTCCAGGTTCCACACAGCGCGGCTCACGGTAACCGATGACAAAGTGTCCTCTA	2358	
Qy	69	GluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThr	88	
Db	2359	AATACCCAAGCTCCCCTCTCGACCAACGCGCGGCCCTCTCCCTACTCTCGSGTCCCCAGC	2418	
Qy	89	GlyPheLeuAlaGluGluAlaProaspAsnValLysThrMetLeuArgSerLysGlyTyr	108	
Db	2419	CTCCACTTAGTGTGAGGAGGAACGAGTAACAGTAACACG-----	2457	
Qy	109	PheSerSerLysValSerLeuThrCluLysaspGlyAlaTyThrValHisIleThrPro	128	
Db	2458	-----GGAGCGGGCTCCAAATTAGCAATAACGCTCTGGCGCTAAAAGTAGGTTC	2508	
Qy	129	GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp	148	
Db	2509	GGTATACCGGTAGATGCTCAAACAGCTCGCTGCATCCCTGGGGGCGGTCTA-----	2562	
Qy	149	GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer	168	
Db	2562	-----	-----	2562
Qy	169	AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----	Ala	185
Db	2563	-----GAAAGCAGAGATAATAAACTGCTGTTAAGGCTGGCGCGGACTTACA	2610	
Qy	186	ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro	205	
Db	2611	ATAACTAATCAAGCTCTTACTGTTGCT---ACCGGGAAC---	GGCTTCAGGTCAACCGG	2664

Qy 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
Db 3625 -----GGCACCGTTAACCTTCTGCGCCGAGGACACCAACT 3663
Qy 531 -----GluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
Db 3664 CTCACGACAGAACCCCTCCGAGCCTCCAACTCCCATCTTCAACTGTCCCTATCG--- 3720
Qy 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
Db 3721 -----GAGGCTGGTGTGTCATAACACGCCCTTCTCTCCAACTGGGAGAGCGCATG 3774
Qy 561 AlaAlaAsnPhelyArgMetGlyLeuLysHisGlySerGlyLeuGlyValArg 578
Db 3775 GAAGTAAATCAGCAGGACTTACTTTAAGAGTAGGCTCGGCTTTCGCAATCGT 3828

RESULT 38

US-08-164-292B-17
; Sequence 17, Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIG
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 29310-20021.00

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 677-7000
; TELEFAX: (415) 677-7522
; TELEX: 34-0154

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

; LENGTH: 5100 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

FEATURE:

; NAME/KEY: CDS

; LOCATION: 408..1331

US-08-164-292B-17

Alignment Scores:

Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Gaps: 36

US-09-857-669-2 (1-609) x US-08-164-292B-17 (1-5100)

Qy 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
Db 2131 ATTAACCCCGCGGCTTCTGGACACAAACGACCTCGCGTCCGCTCTCTTCCCACCGCTG 2190
Qy 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
Db 2191 GCCTCCGATGAGCGCGGACAGCTCAGCTCAACATGTCTGACGGCTATATATAAGGAC 2250
Qy 39 AsnLysSerProAspThrGluSerValLysLysProLysPhe----- 53
Db 2251 AACAGGCTA-----GCTGTCAAAGTAGGTCGCCGGCTGTCCTCGACTCCCAAT 2298
Qy 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
Db 2299 AATCCTCTCCAGGTCACACAGCGGCTCAGGTAACCGATGACAAGGTCTCTCTA 2358
Qy 69 GluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThr 88
Db 2359 AATACCCAAAGTCCCTCTCGACCACGACGCGGCGCTCTCCCTACTTCTGGTCCCAGC 2418
Qy 89 GlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 2419 CTCACATTAGGTGAGGAGGAACGACTAACAGTAAACACC----- 2457
Qy 109 PheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 2458 -----GGAGCGGCGCTCCAAATTAGCAATAACGCTCGCGCTGTAAGTAGGTCTCA 2508
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 2509 GGTATCACCGTAGATGCTCAAAACACGCTCGCTGCATCCCTGGGGAGCGTCTA----- 2562
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 2562 ----- 2562
Qy 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----Ala 185
Db 2563 -----GAAGCAGAGATATAAAACTGCTTAAAGCTGGCGCGGAGCTTACA 2610
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 2611 ATAATAATCAAGCTCTTACTGTGTGCT---ACCGGGAAC---GGCTTCAGTCAACCG 2664
Qy 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 2665 GAA---GGGCAACTGCACCTAAACATTACTCGCGTCAAGGCGCTCAACTTTCACAAACAC 2721
Qy 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 2722 AGCTCGCGTG-----GAGCTGGGCTCGGCGCTG---CAT 2754
Qy 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu----- 256
Db 2755 TTTCCCGCTGGCCAAACCAAGTAAAGCTTTATCCCGAGATGGAATAGACATCCGAGAT 2814
Qy 257 -----LeuLeuAspPheGlnAlaLeu 264
Db 2815 AATAGGTGACTGTCCCGCTGGCGGCGGCTGAGATGCTCAACACCACTTGGCGCTA 2874
Qy 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheArgLeuGlnGly 284
Db 2875 GCT-----TCCGAGACGGTTTAGAGTCCACAGCGACACCCCTC----- 2913
Qy 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
Db 2914 ---CGTTAAAGCTCTCCCGCGCTGACATTTGAAAATGCGCGCTACGACGACAAACTA 2970
Qy 302 GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
Db 2971 GGACGAGACTTGGCACAGACGACTCTGGTCCGCTCGGTTCGACACAGGTTCGAGGACTT 3030

```

/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 345 California Street
/ City: San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94104-2675
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/164,292B
/ FILING DATE: 09-DEC-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: GRACEY, NANCY J.
/ REGISTRATION NUMBER: 28,216
/ REFERENCE/DOCKET NUMBER: 29310-20021.00
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 677-7000
/ TELEFAX: (415) 677-7522
/ TELEX: 34-0154
/ INFORMATION FOR SEQ ID NO: 19:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5100 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 529..954
/ US-08-164-292B-19

Alignment Scores:
Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Gaps: 36

US-09-857-669-2 (1-609) x US-08-164-292B-19 (1-5100)
Qy 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2131 ATTAATACCGCGGGGCTTCTGGAGACAAACGACCTCGCGTCCGCTGCTTTCCACCGCTG 2190
Qy 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2191 GCCTCCGATGAGCGCGGCAACGTACCGTCAACATGCTGACGGGCTATATACTAAGGAC 2250
Qy 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2251 AACAAAGCTA-----GCTGCAAGTAGGTCCCGGGCTGCTCCCTCGACTCCAAT 2298
Qy 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2299 AATGCTCTCCAGTCCACACACGCGGCTCACGGGTAACCGATGACAAGGTGCTCTCTA 2358
Qy 69 GluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLysGluGlnThr 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2359 AATACCAAGCTCCCTCTGACACACACGCGGCTCTCCCTACTCTTCTGGGTCGCCAGC 2418
Qy 89 GlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2419 CTCCACTTAGGTGAGGAGGACGACTAACAGTAAACACC----- 2457
Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2458 -----GGAGCGGGCTCCAAATTAGCAATACGCTCTGGCGGCTAAAAGTAGGTTC 2508

```

Qy	129	GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAlaLeuSerAsp	148
Db	2509	GGTATCACCGCTAGATGCTCAAAACACGAGTCGCTGCATCCCTGGGGACGGTCTA-	2562
Qy	149	GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer	168
Db	2562	-----	2562
Qy	169	AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-	185
Db	2563	-----GAAAGCAGACATATAAAACTGCTGTAGGCTGGCCGGCAGCTTACA	2610
Qy	186	ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro	205
Db	2611	ATAACTAATCAAGCTCTTACTGTGCT--ACCGGGAAC--GGCTTCAGGTCAACCCG	2664
Qy	206	AspThrAlaThrValAspLeuAsnValValValaspSerGlyArgProIleAlaPheGly	225
Db	2665	GAA---GGCAACTGCAGCTAAACATTACTGCGGTGAGGOCCTCAACTTTGCAACAAC	2721
Qy	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
Db	2722	AGCTCGCCGTG-----GAGCTGGGCTCGGCGCTG---CAT	2754
Qy	246	PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu-	256
Db	2755	TTTCCCCCTGGCCAAACCAAGTAAGCCTTTATCCGGAGATGGAATAGACATCCGAGAT	2814
Qy	257	-----LeuLeuAspPheGlnGlnAlaLeu	264
Db	2815	ATAGGGTGACTGTGCCCTGGCGCAGCGCTGAGATGCTCAACCAACCAACTTGCCGTA	2874
Qy	265	GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly	284
Db	2875	GCT-----TCCGGAGACGGTTTAGAAGTCCACACGACACCCCTC-----	2913
Qy	285	AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu	301
Db	2914	---CGGTTAAAGCTCTCCACGCGCTGACATTTGAAATATGGCGCGTACGAGCAAACTA	2970
Qy	302	GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly	315
Db	2971	GGACCAAGGACTGGCACACAGACACTCTGGTCGTCCGTGGTTCACAGGTCCAGGACTT	3030
Qy	316	LysIleAlaTyrAspTyrTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal	333
Db	3031	AGAGTTGCAAACGCCCAAGTCCAGATCTTCAGCGGAAGACGCCCGCATCGGCRCCT---	3087
Qy	334	ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn	353
Db	3088	-----GATAGCAGCTCACTCTCAACATCCGGGCGCCCTCAAA	3126
Qy	354	TyrArgLysAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrTrpGlnAsnLeu	373
Db	3127	TTTTCTGGACCCGCCCTTGACTGCTAGTTTGAAGCAGTGGTCCGATTACTTACAACAGC	3186
Qy	374	GluLysArgAlaPhe-----SerGlyGlyIleTrpTyrValArgAspArgAla	389
Db	3187	AACAATGGCACTTTCCGGTCTCTCTATAGCCCGCGGAATGTGGGTAGACAAAACAGACTT	3246
Qy	390	GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAla	401
Db	3247	CAGGTAACCCAGGCGCTGGTTTACTCTCCAGGAACAACCTTGTCCCAACCTTGGC	3306
Qy	402	GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer	416
Db	3307	GATCCCGCTGGCTATTTCCGACAGCAAAAATTAGTCTCTAGCTCTCGCTCCGCGCTGACCCAA	3366
Qy	417	HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro	436
Db	3367	GCTTCCACGCCCTGACTTTAGT-----TTAGGAACACGGCTTGAATTC	3411
Qy	437	GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer	456

```

Db 3412 TCCAAT---CAAGCGGTGCTATAAAGCGGCGCGGCTTA---CGCTTTGAGTCTTCC 3465
      |||      :::      |||      |||      |||      |||      |||      |||
Qy 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 475
      ::      ::      ::      ::      ::      ::      ::      ::
Db 3466 TCACAAGCTTTAGAGACGACCTCACAGTCGGAAATGGCTTAACG----- 3510
      |||      ::|||      |||      |||      |||      |||      |||
Qy 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn--- 494
      |||      ::|||      |||      |||      |||      |||      |||
Db 3511 CTTACCATACTGTGATCGGCCCAACCTAGGGGACGGCCTAGAGGTCAGACAGCAATAAA 3570
      |||      ::|||      |||      |||      |||      |||      |||
Qy 495 -----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArg 512
      ::      ::      ::      ::      ::      ::      ::      ::
Db 3571 ATCATTTGTAAGTCGGCGGAATCTTCGTTTTTGAACGGGCGGTAAACGCC----- 3624
      |||      ::      |||      |||      |||      |||      |||
Qy 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
      |||      |||      |||      |||      |||      |||      |||
Db 3625 -----GluArgAlaLeuValGlySerLeuGlyTyrGlnLeuProPheThrArg 547
      |||      |||      |||      |||      |||      |||      |||
Qy 531 -----GluArgAlaLeuValGlySerLeuGlyTyrGlnLeuProPheThrArg 547
      |||      |||      |||      |||      |||      |||      |||
Db 3664 CTACTCCAGAACCCCTCCGAGGCTCCAACTCCCATCTTCAACTCGGAGACGGCATG 3720
      |||      |||      |||      |||      |||      |||      |||
Qy 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
      |||      |||      |||      |||      |||      |||      |||
Db 3721 -----GAGGGCTTGTTGTCATAACAAGCGCCTTGCTCTCCAAC TGGGAGACGGCATG 3774
      |||      |||      |||      |||      |||      |||      |||
Qy 561 AlaAlaAsnPhelyArgMetLysLeuLysHisGlySerGlyLeuGlyValArg 578
      |||      ::      |||      |||      |||      |||      |||
Db 3775 GAAGTAAATCAGCAGCGACTTACTTTAAGAGTAGGCTCGGGTTTGCAAAATCGT 3828
      |||      ::      |||      |||      |||      |||      |||

RESULT 40
US-08-164-292B-21
: Sequence 21, Application US/08164292B
: Patent No. 5820868
: GENERAL INFORMATION:
: APPLICANT: MITTAL, SURESH K.
: APPLICANT: GRAHAM, FRANK L.
: APPLICANT: PREVEC, LUDVIK
: APPLICANT: BABIUK, LORNE A.
: TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
: TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 345 California Street
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94104-2675
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/164,292B
: FILING DATE: 09-DEC-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: GRACEY, NANCY J.
: REGISTRATION NUMBER: 28,216
: REFERENCE/DOCKET NUMBER: 29310-20021.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 677-7000
: TELEFAX: (415) 677-7522
: TELEX: 34-0154
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5100 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear

```



```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1246...1707
; US-08-164-292B-21
Alignment Scores:
Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Caps: 36
US-09-857-669-2 (1-609) x US-08-164-292B-21 (1-5100)
Qy 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
Db 2131 ATTAACCCCGGGGCTTCTGGAGACAAACGACCTCGCTCGCTGCTTCCACCGCTG 2190
Qy 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
Db 2191 GCCTCCGATGAGCGCGGACGCTCAGCTCAACATGCTGACGGGCTATATATAAGGAC 2250
Qy 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
Db 2251 AACAAAGCTA-----GCTGTCAAAGTAGTCCCGGGCTGCTCCCTCGACTCCAAT 2298
Qy 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
Db 2299 AATGCTCTCCAGGTCCACACAGCGGCGGCTCACGGTAACGATGACAGGTGCTCTA 2358
Qy 69 GluGluHisLeuProIleThrGlnGlnGluGluValLeuAspLysGluGlnThr 88
Db 2359 AATACCAAGCTCCCTCTCCAGCACCGCGGCGGCTCTCCCTACTCTCGGTCGCCAGC 2418
Qy 89 GlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 2419 CTCCACTTAGGTGAGGAGGAGGACTAACAGTAAACACC----- 2457
Qy 109 PheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 2458 -----GGAGGGGCGCTCCAATTAGCAATAACGCTCTGGCGGTAAGAGTAGGTCA 2508
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 2509 GGATACCGTAGATGCTCAAAACAGCTCCGCTGCTCCCTGGGGGGGGCTA----- 2562
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
Db 2562 ----- 2562
Qy 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----Ala 185
Db 2563 -----GAAAGCAGAGATAATAAAACTGCTGTTAAGGCTGGGCGCGGACTTACA 2610
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 2611 ATAATAATCAAGCTCTTACTGTGCT---ACCGGGAAC-----GGCTTTCAGCTCAACCCG 2664
Qy 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 2665 GAA---GGCAACTCCAGCTAAACATTACTCCCGGTACAGGCGCTCAACTTTGCCAACAAAC 2721
Qy 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 2722 AGCCTCGCGGTG-----GAGCTGGGCTCGGCGCTG---CAT 2754
Qy 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu----- 256
Db 2755 TTTCCCTGGCCAAACAAAGTAAGCCTTTATCCCGGAGTGAATAAGATCATCGAGAT 2814
Qy 257 -----LeuLeuAspPheGlnAlaLeu 264
Db 2815 AATAGGCTGACTGTCGCGCTGGCCAGGCTGAGAATGCTCAACCACTTGGCGTA 2874
Qy 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
Db 2875 GCT-----TCCGGAGACGGTTTAGAAGTCCACAGCACACCTC----- 2913
Qy 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
Db 2914 ---CGGTAAAGCTCTCCACGGCTGACATTTGAAATGCGCGGTACGAGCAAACTA 2970
Qy 302 GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
Db 2971 GGACCAAGACTTGGCACAGACGACTCTGCTGGTCCGTTGTCACAGGTCGAGGACTT 3030
Qy 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal 333
Db 3031 AGAGTTGCAACAGCCCAAGTCCAGATCTTCAGCGGAAGAGACCGCATCGGACT--- 3087
Qy 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
Db 3088 -----GATAGCAGCTCACTCTCAACATCCGGCGGCCCTTACAA 3126
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu 373
Db 3127 TTTTCTGCAGCCCGCTTGACTGCTAGTTGCAAGCAGCTGCTCCGATTACTTACAAACAGC 3186
Qy 374 GluLysArgAlaPhe-----SerGlyGlyIleTyrTyrValArgAspArgAla 389
Db 3187 AACAAATGCGACTTTTGGTCTCTATAGGCCCCGGAATGTGGGTAGACCAAAACAGACTT 3246
Qy 390 GlyIleAspAlaArgLeuGlyValAlaGluPhe-----LeuAla 401
Db 3247 CAGTAACCCAGCGGCTGGTTTGTCTTCCAGGAACAAACCTTGTGCCAAACCTTCGCG 3306
Qy 402 GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer 416
Db 3307 GATCCGCTGGCTATTTCCGACAGCAAAATAGTCTCAGTCTCGGTCCGGCTGACCCAA 3366
Qy 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
Db 3367 GCTTCCACAGCCCTGACTTTAAGT-----TTAGGAACCGGCTTGAATTC 3411
Qy 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 3412 TCCAAAT---CAAGCGGTGCTATAAAGCGGCGCGGGGCTTA---CGTTTGTAGTCTTCC 3465
Qy 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 3466 TCACAAGCTTTAGAGAGCAGCCTCACAGTCGGAATGCTTAACG----- 3510
Qy 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn--- 494
Db 3511 CTTACCGTAGTGTGATCGGCCCAACCTAGGGAGCGGCTAGAGGTGACAGACATAA 3570
Qy 495 -----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArg 512
Db 3571 ATCATTTGTAAGCTGGGCGGGAATCTTCGTTTGAANAACGGAGCGGTAAACGCC----- 3624
Qy 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
Db 3625 -----GGCACCGTTTAACCCCTTCTGCGCGCGGAGGAGCACCACCACT 3663
Qy 531 -----GluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
Db 3664 CTCAGTCGAGAACCAACCCCTCCGAGCTCCCAACTCCCACTCTTCCCTATCG--- 3720
Qy 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
Db 3721 -----GAGGCTTGGTTGTGCATAAACACGCCCTTGTCTCCAACTGGGAGACGGCATG 3774
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArg 578
```

Db 3775 GAAGTAAATCAGACGGACTTACTTTTAAGACTAGGCTCGGGCTTTGCCAAATCGGT 3828

Search completed: November 9, 2002, 02:58:46
Job time : 226 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 02:46:30 ; Search time 61 Seconds

(without alignments)

3542.038 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSDKKIRWHISLGRF 609

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Fgapop 10.0, Fgapext 0.5

Delop 6.0, Delext 7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPro_spool/US09857669/runat_05112002_110010_7843/app_query.fasta_1.775
-DB=PublishedApplications_NA -OFMT=fastp -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09857669 -CGCN_1_1_21_runat_05112002_110010_7843
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	4.6	10419	US-10-027-806-3	Sequence 3, Appli
2	146	4.6	42432	US-10-027-806-2	Sequence 2, Appli
3	129.5	4.1	1398	US-09-938-842A-2084	Sequence 2084, Ap
4	124.5	3.9	2448	US-09-815-242-4051	Sequence 4051, Ap

5	122	3.8	6026	9	US-09-995-749A-1	Sequence 1, Appli
6	121.5	3.8	640681	10	US-09-790-988-1	Sequence 1, Appli
7	116	3.7	2950	10	US-09-886-468-6	Sequence 6, Appli
8	112	3.5	2439	10	US-09-974-300-472	Sequence 472, App
9	112	3.5	3213	10	US-09-974-300-475	Sequence 475, App
10	112	3.5	3286	10	US-09-332-226-1	Sequence 1, Appli
11	111	3.5	34446	10	US-09-871-212-1	Sequence 1, Appli
c 12	108.5	3.4	1969	10	US-09-864-761-4449	Sequence 4449, Ap
c 13	108.5	3.4	2004	10	US-09-864-761-21197	Sequence 21197, A
14	108.5	3.4	4896	10	US-09-740-274-3	Sequence 3, Appli
15	108.5	3.4	9477	10	US-09-815-242-8513	Sequence 8513, Ap
16	108	3.4	4065	12	US-10-044-090-98	Sequence 98, Appli
17	107.5	3.4	14800	10	US-09-954-456-1601	Sequence 1601, Ap
18	106.5	3.4	2140	10	US-09-974-300-2290	Sequence 2290, Ap
19	106.5	3.4	2140	10	US-09-974-300-2620	Sequence 2620, Ap
20	106.5	3.4	3537	10	US-09-332-226-3	Sequence 3, Appli
21	104.5	3.3	1149	10	US-09-912-020-139	Sequence 139, App
22	104	3.3	2055	10	US-09-974-300-2362	Sequence 2362, Ap
23	103.5	3.3	4557	10	US-09-974-300-2632	Sequence 2632, App
24	103.5	3.3	5480	10	US-09-070-927A-132	Sequence 132, App
25	103	3.2	1164	10	US-09-974-300-28	Sequence 28, Appli
26	102.5	3.2	1389	9	US-09-938-842A-2606	Sequence 2606, Ap
27	102.5	3.2	22671	10	US-09-956-004-14	Sequence 14, Appli
28	101.5	3.2	15231	10	US-09-917-800A-1505	Sequence 1505, Ap
29	101	3.2	2949	10	US-09-841-132-170	Sequence 170, App
30	101	3.2	3021	10	US-09-841-132-182	Sequence 182, App
31	101	3.2	7107	10	US-09-070-927A-527	Sequence 527, App
32	100.5	3.2	2820	10	US-09-815-242-6700	Sequence 6700, Ap
c 33	100.5	3.2	13884	10	US-09-070-927A-341	Sequence 341, App
34	100.5	3.2	32768	10	US-09-070-927A-17	Sequence 17, Appli
35	100	3.2	6258	10	US-09-815-242-4584	Sequence 4584, Ap
36	100	3.2	17388	10	US-09-815-242-8512	Sequence 8512, Ap
37	100	3.2	35100	10	US-09-782-378A-26	Sequence 26, Appli
38	99.5	3.1	2085	10	US-09-974-300-7800	Sequence 7800, Ap
39	99.5	3.1	3078	9	US-09-938-842A-1039	Sequence 1039, Ap
40	99.5	3.1	3412	10	US-09-897-056-6	Sequence 6, Appli
41	99.5	3.1	6077	10	US-09-815-242-4648	Sequence 4648, Ap
42	99.5	3.1	6114	9	US-09-938-842A-2700	Sequence 2700, Ap
c 43	99	3.1	1243	10	US-09-070-927A-504	Sequence 504, App
44	99	3.1	1800	8	US-08-973-028-3	Sequence 3, Appli
45	99	3.1	2295	10	US-09-747-521-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-027-806-3
; Sequence 3, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 10419
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(10419)
US-10-027-806-3

Alignment Scores: 9.28e-06 Length: 10419
Pred. No.: 128
Score: 146.00 Matches:

ORGANISM: Cenarchaeum symbiosum

FEATURE:
 NAME/KEY: CDS
 LOCATION: (3)...(10421)
 NAME/KEY: CDS
 LOCATION: (10625)...(11434)
 NAME/KEY: CDS
 LOCATION: (11478)...(13046)
 NAME/KEY: CDS
 LOCATION: (13046)...(14620)
 NAME/KEY: CDS
 LOCATION: (23558)...(24862)
 NAME/KEY: CDS
 LOCATION: (24913)...(25728)
 NAME/KEY: CDS
 LOCATION: (26504)...(26881)
 NAME/KEY: CDS
 LOCATION: (29655)...(30491)
 NAME/KEY: CDS
 LOCATION: (34559)...(36067)
 NAME/KEY: CDS
 LOCATION: (37002)...(37403)
 NAME/KEY: CDS
 LOCATION: (37404)...(38282)
 NAME/KEY: CDS
 LOCATION: (39454)...(40572)
 US-10-027-806-2

Alignment Scores:

Pred. No.: 8,07e-05 Length: 42432
 Score: 146.00 Matches: 128
 Percent Similarity: 31.17% Conservative: 59
 Best Local Similarity: 21.33% Mismatches: 213
 Query Match: 4.60% Indels: 200
 DB: 9 Gaps: 31

US-09-857-669-2 (1-609) x US-10-027-806-2 (1-42432)

QY 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
 DB 6261 GGCAGCAGCGCTGAATCCGCGTAATATCCACGAGGGCTCGACACCTGTACAGCTTT 6320
 QY 158 AlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
 DB 6321 GTACTGGAC-----ATCCGCGTATGGCGGAATTGGATATTCACAGGCTTGAG--- 6368
 QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
 DB 6369 -----CTTCCGCTGGTGGGGTTCCGACG-----GGATTCCAGTTCTCG----- 6407
 QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
 DB 6408 -----GAC 6410
 QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 DB 6411 AACGGCGCCAGTTGTACATTTGGCGCTTTCGT-----GACTCTCAATCTCCGCCAGGC 6464
 QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu--- 256
 DB 6465 ACCCTGCTCGGGCCCTGCAGCGCTATGAGCTTGGCATACCATATGACCTGGCTTCGGCT 6524
 QY 257 -----LeuLeuAspPheGln-----GlnAlaLeuGluGlnAsn 267
 DB 6525 GTATTTGGCGAGTCCCTGGGAATATTTCGATTTTCTCCCTTCAACGGCATGCGGGCCAAT 6584
 QY 268 GlyHisThrSerGlyAlaSerValGlnAlaAsp-----PheAspArgLeuGln 283
 DB 6585 GGCAGCTTGGCAGGATTACATGTGCCCGCGATGGAGCATCCTGTCAGGGCGGGAAT 6644
 QY 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
 DB 6645 GCGGAAGAACCGTAATCAGCTAT-----GACATGGACAGCCATGATTTGGATACA 6695

QY 304 -----GlyIleArgLeu 307
 DB 6696 TTATCATTCAGGGAATCATTCAAACACAGATGCGGACAGTCGCACACCCACATAAGS--- 6752
 QY 308 AspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr-----Asp 320
 DB 6753 GACATGGACATATCCCGGACGGCATGTTCTCTACCTGCTCAAGGCGATGTTCTGGAC 6812
 QY 321 TyrTyrAsnLeuPheAsnLys-----GlyTyrIleGlySerValVal 334
 DB 6813 ATGTACAACCTTACAGATAGTATTTCGTTGATGCCCGCGCATATCGGGTACCTG--- 6869
 QY 335 TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr 354
 DB 6870 ---GATTGGAAACCG---GAGGATGTAATACCCAGGGGATTTTCATCTCACCAGGATGGC 6923
 QY 355 ArgGlyAsnTyrTrpThrSer----- 361
 DB 6924 ACCAGTCTGTTTATGACAGCGGAAGACGTGGACCACATTTCACCAATATGCATTGAATGAA 6983
 QY 361 ----- 361
 DB 6984 CCATGGGACATAGCAATGCCATPACTTGCAGGCTCCCTGTCCATAAGCCAGTGAATGGT 7043
 QY 362 -----AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
 DB 7044 GCACCCCGGGGGCTGGATATATCGGAGGATGGCAACACTGCACATCTACTATGCCGGCGCT 7103
 QY 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIle----- 391
 DB 7104 GACTTTTGACACGGGG-----CCCGCATCCCTGGTAAACACCATATTTG 7145
 QY 392 -----AspAlaArgLeuGlyAlaGluPheLeuAlaGluGly 403
 DB 7146 CCAGGCCAATATTCCTGCTGACGGATGCGCGCGCTTTGCA---TACCCTGGTGGAGGAG 7202
 QY 404 ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
 DB 7203 GAGGTGTCACCGGG---GATCTTGCACTTCCGATGACGGCATGGCATGTTCTGGTGGCG 7259
 QY 424 SerTrpLysArgGlnLeuLeuAsn---AsnValLeuHisPro-----GluAsnGly 439
 DB 7260 GGCCTAAACAAACCATTTAAGACAGATACAACCTGCTGTCGCGCTATGACACTGAAATATGCA 7319
 QY 440 ---HisTyrLeuAspGlyLysIle-----GlyThrThrLeuGlyThrPhe 453
 DB 7320 GAACATTTTCATCTCGACGGATCTGCTGCTGACGACAGGGGGCCCGCATGTTGATATT 7379
 QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr-----PhePhe 470
 DB 7380 TCAGATGAGAACCACTTTTTCACGACAGCGCCAGGGCCCAATTTGTGCCCGCATTTACG 7439
 QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
 DB 7440 ACAACCGCCCGTACGAGCATCCACAATAACACTGAGTCACACGAGCTGTACAAGGTG 7499
 QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
 DB 7500 AGCGTGGACGGC-----CTGCGCTCGGCATACGGTTTACCCCGCAGCGCATGAAGATG 7553
 QY 511 ---ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeu 529
 DB 7554 TTCATATCGGCGGACGAGCGGCCCATGATA----- 7583
 QY 530 ProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeu 549
 DB 7584 -----TACCAGTATTCCTCGCGTCCCGCATGACACA 7616
 QY 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
 DB 7617 TCCGGGCGGCTCAGGACAGGGTTGAGATAGTGCAGGGCTCTTTAGAAATGCAGGTTTG 7676

QY 570 Lys----- 570
Db 7677 TCCGTCGGTGAACGAGCCAGCTCTCCGGCTTCACTTTTCGAGGACGAATGGAG 7736
QY 571 -----HisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSer 587
Db 7737 CTCTAGCTGACGGGTGGGCGCTGTTCACAGGATATTCCTGCCA---TCGCCATACGGC 7793
QY 588 Phe---AspIleAlaTrpGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGly 606
Db 7794 CTCGAAGATGACAGCTACGGGGCGAGCTTCCACACGTTTCAGGGAGAGACGCGCGTGGGA 7853
RESULT 3
US-09-938-842A-2084
; Sequence 2084, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2084
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2084
Alignment Scores:
Pred. No.: 3.16e-05 Length: 1398
Score: 129.50 Matches: 121
Percent Similarity: 35.50% Conservative: 54
Best Local Similarity: 24.54% Mismatches: 198
Query Match: 4.08% Indels: 121
DB: 9 Gaps: 26
US-09-857-669-2 (1-609) x US-09-938-842A-2084 (1-1398)
QY 96 ProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu 115
Db 103 CCTTCGATTCGAAACT-----CGTAGCTTAAGAAACCATCTCTGTACGAGCTGGCTTA 156
QY 116 ThrGluLysAspGlyAlaTrpValHis---IleThrProGlyProArgThrLysIle 134
Db 157 ATCGAGCCAGATGGTGGAAACTTGTGGATCTTGTGTACCGGAACCGAGACGGCGAGAG 216
QY 135 AlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyr 154
Db 217 AAGAAACACGAGCGGGGATTTCGCGAGAGTGGAGATTCGCGCGATTCGATTCGAATGG 276
QY 155 TyrArgAsnAlaLeuGluAsnTrpGlnInProVal-----GlySerAsp 169
Db 277 ATGCATCTGTGAGTAGAGGTGGCTAGTCCTCTCGTGGTATTATGAGGGAACTCTGAG 336
QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
Db 337 TTCCTCCAAACTCTTCATTTTCAATTTGTTGAATCTCGATGGGTCTGTGTTTAATG 396
QY 190 GlyTyrProLeuAla---LysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla 208
Db 397 TCTGTGCTTATTTCTTCGCTATGATGATCAACAAACAAACCCCTAATTCGCTGAATCTAAA 456

QY 209 ThrValAspLeuAsnValValValAspSer---GlyArgProIleAla---PheGlyAsp 226
Db 457 CGTGTCTCCCTT-----GTTGATTCTGATGATAATCAATCCATATTCATTCATGAT 507
QY 227 PheGluIleThrGlyThrGlnArgTyrPro-----GluGlnIleValSerGlyLeuAla 244
Db 508 ATTGAGATT-----TATAAACATCCGAAAGAGAGGAATAGCGAGAACCTGGGGT 558
QY 245 ArgPheGlnProGlyThrProTyr----- 252
Db 559 ACGACTGCACCGGTTTGCCTTATGTAGAAGAGCGGATAACCAATGCTGGAGACTGGCTC 618
QY 253 -----AspLeuAspLeuLeu----- 257
Db 619 ATTGGGGGTGATCTTTCAGGTTTGGAACTGTTAAGTACAAATGATGGCTTGCCTTTC 678
QY 258 -----LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 679 AGGCTTTCCCGGTTTGAACCTGCGTAAGGAGCTAGAGAAACGTGTCGGGATCGGCTCTTT 738
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProVal----- 289
Db 739 GCCTTTCAGCTT-----AGGAACCCAGTTCATAATGGACAT 774
QY 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg-----Leu 307
Db 775 GCTCTTCTTATGACTGATCTCGTAGGACTTCTTGAGATGGGTTATATAAACCCCTATC 834
QY 308 AspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys 327
Db 835 CTTTTTGTTCATCCATTGGGAGG-----TTTACAAAA 867
QY 328 GlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaIleGly 347
Db 868 GCTGATGATCTCCTCTAAGCTGGCGAATGAACACGACGAGAGGTGCTGAGGATGGT 927
QY 348 IleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg 367
Db 928 GTTCTTCATCCAGAGACTACTGTGTTTCCATATTCCTCATCTCCATGCTCTATGCTGGT 987
QY 368 SerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTyrTrpTyrValArgAsp 387
Db 988 CCAACCGAAGTCCAA-----TGGCAGCGAAAGGCT 1017
QY 388 ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIlePro 407
Db 1018 AGG-----ATCAATGCT-----GGTGTAACTTCTACATTCGCGTAGGGATCCGGCT 1065
QY 408 GlySerAsp-IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla----- 423
Db 1066 GGAATGGGTATCCCGTGGAGAAACGCTGATCTGTACGATCTGATCACGGGAGAAAGTT 1125
QY 424 ----SerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLe 442
Db 1126 CTAAGCATGG-----CTCCT 1140
QY 442 uAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgTh 462
Db 1141 GGACTCGAAGCACTCAACA-----TTCCTCTTTTCA-----GGGTTCGTCGTAC 1185
QY 462 rSerAlaArgAlaGly-----TyrPhePheThrProGluAsnLysLysLeuGlyThrPh 480
Db 1186 GAT-AAGACACAAGGAAGATGGCTTTTGTATCCCTCAAGGGCTCAGGACTTCTTGT 1244
QY 480 eIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerG 500
Db 1245 C---ATTTCGGCACTAAGATGAGAGCATTTGGCAAGAACAGAGAGAACCCACAGATGG 1301
QY 500 yLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleG 520
Db 1302 ATTTATGTGCCCTGGAGCGTGGAAAGTCTTGTGGATTACTAC-----GACAGTTTGAC 1355
QY 520 yLeuAlaGlyProAsnGlySerValLeuProGluArg 532

```
Db 1356 ACTAACCGGA-----AACACCAAACTTCGGGAAAG 1386
RESULT 4
US-09-815-242-4051
; Sequence 4051, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4051
; LENGTH: 2448
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4051

Alignment Scores:
Pred. No.: 0.000277 Length: 2448
Score: 124.50 Matches: 140
Percent Similarity: 33.91% Conservative: 77
Best Local Similarity: 21.88% Mismatches: 227
Query Match: 3.92% Indels: 198
DB: 10 Gaps: 34

US-09-857-669-2 (1-609) x US-09-815-242-4051 (1-2448)
Qy 66 AspMetValGluHisLeuProLeuThrGlnGlnGluValLeuAspLys 85
Db 556 GACGACGTCTACGCCATACCGCGGCATCACC-----GTCCTCGGCTACGACACT 606
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaPro-----AspAsnVal 99
Db 607 GACCGCAACAACACTACTATGCGCGCGCTTCGATCAACAACATTCACGACGGGATT 666
Qy 100 LysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal-----SerLeuThrGlu 117
Db 667 CCCTCGAGCGCCGACAGTCGGCTACTCCGCGGGCAATAGCTCAGTACGATCGCCATC 726
Qy 118 LysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137
Db 727 TACGACCGGGTGAAGTACTCAAGGGGCGACCGCTGCTCACCGGTGGGGTCCCTC 786
Qy 138 GlyValAlaIleLeuGlyAspIleLeu-SerAspGlyAsnLeuAlaGluTyrTyrArgAs 157
Db 787 GGGCCCAACGATC-----AACCTGATCGCAAGAAACCTACCCATGAATTAAGG-- 835

157 nAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAs 177
836 -GCCATGTCGAACCTGGCGCGGCGAGTTGGACA--ATTACCGCTCCGAACCTGGACGTC 891
Qy 177 nSer-----LysThrSerValLeuGlyAla----- 185
892 AGCGGCCCACTGACCGAGTCCGGCAACGTCGCGCGCGCGGTAGCGGCTACCAAGAT 951
186 -----ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrAr 200
952 AAGCACTCGTTTCATGGATCACTAGCAGCGCAAAACACAGCTCTATTACGGCATCTCGAG 1011
200 gAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAsp----- 217
1012 TTCGAT-CTGAATCCCGACACCATG-----CTTACCGTCGGCGCCGACTATCAGGA 1061
Qy 218 -----SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyTh 232
1062 CAACGATCCCAAGGGATCGGGCTGGTGGCGAGCTTCCCGCTCTTCGACACCGCAGGCAA 1121
232 rGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTy 252
1122 CCGCAGCAGCTCTCCCGCTCTTCAACAGGAGCCAGTGGAGCAGTTGGGAGCAGTA 1191
252 rAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln----- 266
1182 CACCCGCGACGCTCTTCCCAACCTCGAGCACAATTTCCGCCACCGCTGGGTGGCAAGGT 1241
Qy 267 -----AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAs 280
1242 GCAACTCGATCACAAAGATCAACGGCTACCAC-----CGCGCCCTCGG 1283
280 pArgLeuGlnGlyAsp-----ArgValProValLysValSe 292
1284 CGCATCATGGCGGACTGGCGCGGACCGGACACAGCCGCAAGATCTTGGCAGAGTA 1343
292 rValThrGluValLysArgHisLysLeuGlu-----ThrGly----- 305
1344 CACCGCGGAAACCAAGAGCAACTCGCTGGATATCTATCTACCGCGCCCTTCCAGTTCT 1403
305 eArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPh 325
1404 CGGTGCGGAGCAGCAGCTGGTGGTGGCGACCTCGCTTCTTCTCCCACTGGGAG----- 1458
325 eAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThrThrLeuAl 345
1459 -----GGCAAGAGCTACTGGACCTCGCAACCTGGCAACACACACACCGCA 1502
345 aAlaGlyIleSer-----GlnProArgAsnTy 354
1503 CGACTTCATCACTGGGCGGCGGATATAGGCAACCGCGCTGGGTACCCCTCCCAAGTA 1562
354 r-----ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnAr 367
1563 CATCGATGACAAGACCGCGGCTGGGCGAGCTACATGACCGCGCG-----TTCAATGT 1616
367 gSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleThrTyrValArgAs 387
1617 GACCGATGACCTGAACCTG-----TTCCTAGGTGGA----- 1647
387 pArgAlaGlyIleAspAlaArg-----LeuGlyAlaGluPheLeuAlaGluGlyAr 404
1648 -----CGCGTGGTGCATCTACCGGCTAACCGGCTGAATCCGACCATTCGCGAGTCCGACG 1703
404 gLysIlePro-----GlySerAspIleAspLeuGlyAsnSerHisAlaThr----- 419
1704 CTTCAATCCCTACCTCGCGCGGCTCGACCTCAACGATACCTACTCCGCTACGCCAG 1763
420 -----MetLeuThrAlaSerTyrLysArgGlnLeuLeuAsnValLe 434
1764 CTATACCGACATCTTCATGCGCGCAGGACAGTTGTGTGTCGCGACAGCAGCAACAGCTGCT 1823
434 uHisProGlu-----AsnGlyHisTyrLeuAspGlyLy 445
```

```

1      |||:::
Db 1824 CGAACCGCAGCAAGGCCAGAACTACGAGATCGGCATCAAGGGTGAATATCTCGACGGACG 1883
Qy 445 sileGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuLeuLeuLeuLeuLeuLeuLeu 465
Db 1884 ACTCAATACCAAGCTCGCCCTACTTC---GAGATTCAAGAGAAACCGCGGAGGAAGA 1940
Qy 465 gAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly-- 484
Db 1941 CGCCCTCTACAACAGCAACCGACCAACCCCGGCATC---ACCTACGCCTACAAGGGCAT 1997
Qy 484 ----- 484
Db 1998 CAAGGCCAACCAAGGGCTACGAGCGGGAATCTCGGGTGAACCTGCGCGCAGGCTGGCA 2057
Qy 485 ----GlnAlaGlyTyrThr-----ValAlaArgAspAsnAlaAspValProSerGlyLe 501
Db 2058 GGTTCAGGCTGGCTATACCCACAAGATCATTCGGGAGCAC-----ACTGGCAA 2105
Qy 501 uMetPheArgSerGlyGlyAlaSer-----SerValAr 512
Db 2106 GAAGGTTTCCACCTG-GGAGCCTCAGGACCAGTTGAGCCTCTACACAGCTACAAGTTCA 2164
Qy 512 gGlyTyrGluLeuAspSerIlegly-----LeuAlaGly----- 523
Db 2165 AGGCGGCCCTGGACAAGCTCACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2224
Qy 524 -----ProAsnGlySerValLeuProGluArgAlaLeuLe 535
Db 2225 AGATGGTCTACACACATCCCGCAGCGCGTGGGAAAGATTCTCCCGAAGA---CTACT 2281
Qy 535 uValGlySerLeuGlyTyrGlnLeuPro-PheThrArgThrLeuSerGlyAlaVal 553
Db 2282 GGCTGGTGGACCTGATCGCCCGCTACCATCAGATACCGACAAAGCTGTACAGCCGCGTC 2337

```

RESULT 5

```

US-09-995-749A-1
; Sequence 1, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: 8043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 6026
; TYPE: DNA
; ORGANISM: Lactobacillus reuteri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(5503)
US-09-995-749A-1

```

Alignment Scores:

```

Pred. No.: 0.00213 Length: 6026
Score: 124.00 Matches: 128
Percent Similarity: 30.17% Conservative: 85
Best Local Similarity: 18.13% Mismatches: 231
Query Match: 3.84% Indels: 262
DB: 9 Gaps: 34

```

US-09-857-669-2 (1-609) x US-09-995-749A-1 (1-6026)

```

Qy 35 AlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhePro 54
Db 3485 TCATTACGTGATCGTGAAATAATTCACGGAATCTGTAAATT---CCTAATATACAGT 3541
Qy 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
Db 3542 TTC---GTTCCGGGCTCAGATAAATTCACAGATCAAAATTCAAAATGCTAATT----- 3592
Qy 75 IleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGlu 94
Db 3593 -----CCTGATGTAAACAGGCAAGATTACCATACTTTTCACTTTTGAAGAT 3637
Qy 95 AlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 3638 GAGCAAAAGGGTATTGATGCGTACATTCAAGATCAAAATTCACAGTGGAAGAAATATAAC 3697
Qy 115 LeuThrGluLysAspGlyAlaTyrThrValHisIleThrProGly-----ProArg 131
Db 3698 CTTTATAAATATTCGGGCTTCATACGCAATCTTTTAACTAACCAAGGATACAAATCCACGT 3757
Qy 132 ThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly----- 149
Db 3758 -----GTATACTATGGTGACTTGTTATATCTGATGCTGATGCTGGGCCAA 3793
Qy 150 -----AsnLeuAlaGluTyrTyr----- 155
Db 3794 TACATGGAACATCAACACACGTTACTATGATCTTTTAAACGAACCTGCTTAAATCACAGATT 3853
Qy 156 -----ArgAsnAla 158
Db 3854 AAGTATGTTCCCGTGGCCCAATCAATGCAACCAATGAGCGTTGGCGCAATAATAACATT 3913
Qy 159 LeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSer---TrpGluAsn 177
Db 3914 TTACTACTGTTCTGTTATGTTAAAGTGGCGATGACACCTACTGATACCTGCTGACTGCTGAA 3973
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 3974 ACCAGAACAACAGGTATTGGGGTTGTTGTA---AGTAATACGCCAAATCTAAACCTAGGT 4030
Qy 198 -----AsnThrArgAla 201
Db 4031 GTCAACGATAAAGTAGTTCTTTCATATGGGAGCTGCGCACAGAACCACCAATAATPCGGGCA 4090
Qy 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 4091 GCGTGTGTGAGCACAACTGATGGAGTC---ATTAAATTATACTTCTGATCAAGGGCCACCG 4147
Qy 222 IleAlaPhe-----GlyAspPheGluIle----- 229
Db 4148 GTTGCAATGACTGACGAGAACGGTGATCTATCTATCTATCTAGCTATAACCTAGTTGTTAAT 4207
Qy 230 -----ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 4208 GGTAAAGAAGACGACATACAGCTGTTTCAAGGTTATGCTAACCCCTGATGTTTCAAGTAT 4267
Qy 244 -----AlaArgPheGlnProGly 249
Db 4268 CTTGCTGTATGGGTACCAGTTGGGCAAGTGTATACCAAGATGCTCGAACTGCTCCATCT 4327
Qy 250 ThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHis 269
Db 4328 ACT-----GAAAAGATAGTGGT 4345
Qy 270 TyrSerGlyAlaSerValGlnAlaAspPheAspArg-----LeuGlnGlyAspArg 286
Db 4346 AACTCTGCATACAGACAAATGCTGCTTTTGCATTCAATGTTATTTTGAAGCCCTTTTCT 4405
Qy 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg 306
Db 4406 AACTTTGCTCTATACACCAACAAAGGAAGTGAACCTGCTTAATGTTCCAAATGCCCAANT 4465
Qy 307 LeuAspSerGluTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyrTyrAsn 323

```


Db 262199 CAAATTGATATTACCTATCAAGTTAAAGACAGACTACTGGTCTATAATAATTTTGGTTTA 262258
Qy 306 ArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
Db 262259 GGATATGGAGTAGATAGTGGTACAGATTTAATCTTCTGCTTTTCTCAAGATAATATATT 262318
Qy 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLys-----TyrGluThrThr 343
Db 262319 GGTTCCTGGA-----AATTCCTTAAAGTCGATATTATTAATAATGATTATCAAAAATAC 262372
Qy 344 LeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsn----- 357
Db 262373 CTTGATATATCAACAAGTATTCATATTTTATATATATGATGAGATTTAAATGCCAGA 262432
Qy 358 TyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAla 377
Db 262433 TTTTTCACATGATTTTAAATATAATTTTGACATATTTTCGAATATTATAAAAAATACA 262492
Qy 378 PheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 262493 TAT-----GGTTTGAAGCTAATTTGGGA 262516

RESULT 7

US-09-886-468-6
; Sequence 6, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses thereof
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn ver. 2.0

SEQ ID NO 6
LENGTH: 2950
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(2866)
US-09-886-468-6

Alignment Scores:
Pred. No.: 0.00341 Length: 2950
Score: 116.00 Matches: 131
Percent Similarity: 30.23% Conservative: 83

Best Local Similarity: 18.50% Mismatches: 226
Query Match: 3.66% Indels: 268
DB: 10 Gaps: 29
US-09-857-669-2 (1-609) x US-09-886-468-6 (1-2950)
Qy 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
Db 380 TATTTTAATAATATTTCTCAGGAACACAAAGGAGGGGCT-----GTACTTTGTTGC 433
Qy 128 ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSer 147
Db 434 CAAGATCCTCAAGCAACGGCAGCTTTTCTGGGTCTCCACGCTCTTTTATTTATTCAGAC 493
Qy 148 AspGlyAsnLeuAlaGlu-----TyrTyrArgAsnAlaLeu----- 159
Db 494 CCGGAGATATTAAGAACAGGGATGCTCTATTTCAAAAAATGACATTTATGCTCTTAAC 553
Qy 160 -----GluAsnTrpGlnGlnProValGly----- 167
Db 554 AATTATCTAGTGCCTTTTGAACAAACCAAGTAAGACTAAAGCGGAGCTATTAGTGGG 613
Qy 168 -----SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
Db 614 CGCAATCTTACTATAGTAGGCACTACGATTCCTCTCTTTCTATCAGAATGCAGCCACT 673
Qy 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeu----- 193
Db 674 TTTGGAGTCTATCCATTTCTCAGGT---CCCCACAGATTGGCAGTAAATCAGCGCAGAG 730
Qy 194 -----AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
Db 731 ATAAGATTTGCACAAATACTGCCCAAGATGTTCTCGAGGGCTTTGTACTCC----- 784
Qy 207 ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAsp 226
Db 785 -----GATGGTGATATTGATATTCATCAGAAATGCTTATTTATTCGAGAA 832
Qy 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
Db 833 AATGAGGCATTTGCTACTGCTATAGTTAAGGGAGGGCTGCTCTGTTCTCTCCACTTCA 892
Qy 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 893 GGAAGTAGTACTCCAGTTCTATTTGCTCTCACAATAAACAGTAGTCTTTTGA 952
Qy 267 AsnGlyHis-----TyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
Db 953 AGAAACCATTTCCATAAATGGTGGCGGAGCCATTTATCTAGG----- 994
Qy 285 AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGly 304
Db 995 -----AACTTTAGCATCTCTTCAGAGAGTCTCTACTCTATTTATCAATAAT 1039
Qy 305 IleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp----- 320
Db 1040 ATATCATATGCAAAATTCGCAAAATTTAGTGGAGCTATTGCCATTGACTACTGGAGGGAG 1099
Qy 320 ----- 320
Db 1100 ATCAGTTTATCAGCAGAGAAAGAACAAATTACATTCCAAGAAACCGGAGAGCTTACCG 1159
Qy 320 ----- 320
Db 1160 TTTTGAATGGCATCCATCTTTTACAAAATGCTAAATTCCTGAAATTTACAGCGCAGAAAT 1219
Qy 321 -----TyrTyrAsnLeuPhe----- 325
Db 1220 GGATACTCTATAGAAATTTTATGATCCCTATTACTTCTGAAGCAGATGGTCTCTACCAATTG 1279
Qy 326 -----AsnLysGlyTyrIleGlySerValValTrpAspMetAsp 338
Db 1280 AATATCAACGGAGATCTTAAANAATAAGACTACACAGGACCATCTCTTTTCTGGAGAA 1339

```

Qy   593 GlyHisSerAspLysLysLeArg 600
      |||||:::          :::::
Db   2321 GGTCA TGCCACCAATGATGAAA 2344

RESULT 8
US-09-974-300-472
; Sequence 472, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIORITY APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 472
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-472

Alignment Scores:
Pred. No.:           0.00725       Length:        2439
Score:              112.00         Matches:       119
Percent Similarity: 35.11%         Conservative:  72
Best Local Similarity: 21.88%      Mismatches:   187
Query Match:        3.53%          Indels:       166
DB:                  10            Gaps:          29

US-09-857-669-2 (1-609) x US-09-974-300-472 (1-2439)

Qy    20 TyrAlaProAlaAlaSpLeuSerGluAsnLysAlaAlaGlyPhe---AlaLeuPheLys 38
      |||||:::          :::::
Db    966 TATCGGACCCTGCATTTCGGTTCGGGCGATTAAGAAAGCGGCTATGAAGCCATTATTATC 1025

Qy    39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPheProValargIleasp 58
      |||  :::::||:::|||  |||
Db    1026 AACAAACAATCCTGAGACGGTTTCAACA----- 1052

Qy    59 ThrGlnAspSerGluIleLysAspMetValGluHISLeuProLeuIleThrGlnGln 78
      |||  |||  |||  :::::
Db    1053 -----GACTTCAGCATATCCGACAGAAGCTCTACTTTGAG--CCGCTTACTGTTGAAGAT 1103

Qy    79 GlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsn 98
      ::::|||||:::
Db    1104 GTCATGTCATATCATCGATTTCGAGCAG-----CCGAAAGGC 1139

Qy    99 ValLysThrMetLeuArgSerLysGlyTyrrPheSerSerLysValserLeuThrGluLys 118
      |||  |||  :::::
Db    1140 GTCGTCGTTTAG-----TTTGCGGACACACTGCCGATCAACCTTGCC 1181

Qy    119 AspGlyAlaTyrrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly 138
      |||  |||  :::::
Db    1182 GAC-----TTTGCGGACACACTGCCGATCAACCTTGCC 1181

Qy    139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu-----TyrTyr 155
      |||||:::
Db    1203 GTCGCCAFTTCGGAACATCCCTTTGAAGATTTCGCGCGGACGGACGGCACAAAGTTC 1262

Qy    156 ArgAsnAlaLeuGluAsnTrpGln-----GlnProValcylSerAspPheAspGlnAsp 173
      |||||:::
Db    1263 GAACAGACGGCTTGNAATGTCGACGTTACCCACCGCCCTTGG----- 1304

Qy    174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaVal-----ThrArgLysGly 190
      :::  |||||:::

```

```
Qy 191 TyrProLeuAla-----LysLeuGlyAsnThrArgAlaAlaValAsnPro 205
|||||
Db 1356 TATCGGTGCTGTCGCGCCCATCGTATCTACTCGCGGACGTCGATGGAGATCGTCTAC 1415
|||||
Qy 206 AspThrAlaThr-----ValAspLeuAsnValValValAspSerGlyArgPro 221
|||||
Db 1416 CATGAACGACGAGCTCTCCATTATATGAAATGCAATGAAATCAATCCCGACACGCT 1475
|||||
Qy 222 IleAlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSer 241
|||||
Db 1476 GTCTGTGATC---GATAAATATTGACCGGAAAGAAATAGATGGATGCCATTCTTCGAC 1532
|||||
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
|||||
Db 1533 GGTGAACGGTGGCTCATTCGCGGG-----ATTATG 1562
|||||
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
|||||
Db 1563 CAGCACATCGAACGGGGGGGCTCCACTCGGGGATTCGATT----- 1604
|||||
Qy 282 LeuGlnGlyAspArgValProValIleValSerValThrGluValLysArgHisLysLeu 301
|||||
Db 1605 -----GGCGTTTACCCCGCGCAATCCCTTTCAGAGGACATCAAGAAAAGCTT 1652
|||||
Qy 302 GluThr---GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
|||||
Db 1653 GAAGCATACAGATTAAAGTAGCAAGGGCTTAAACATCGTCGCGCTCTCAACATTCAG 1712
|||||
Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyr 340
|||||
Db 1713 TTTGTCCTTTCAAAAGATGAGCTTTACGCTCTTGAAGTC----- 1751
|||||
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
|||||
Db 1752 -----AACCCGAGATCGAGCCGCAAGCGTCCGCTTCTC 1784
|||||
Qy 361 SerAsnValSer-----TyrAsnArgSerThr----- 369
|||||
Db 1785 AGCAAAATCACCGGATTCGATGGCAATCTCGCAACAAAGGTGATTCTCGGAGGGAAG 1844
|||||
Qy 370 -----ThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrp 383
|||||
Db 1845 CTGCCGATATGGGCTATACAGAGGCTCCAAAAGAG-----CAGGAAGGGGTA--- 1895
|||||
Qy 384 TyrValArg-----AspArgAlaGlyIleAspAlaArgLeuGly 396
|||||
Db 1896 TATGTGAAGTCCGGTCTTCTGTTTCCGAAGCTTAGAAGAGTGGATATCACGCTCGGA 1955
|||||
Qy 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp----- 410
|||||
Db 1956 CCGGAATGAAATCGACCGGG---GAAGTCATGGCAAGACAGACACATTCGAAAGAGCG 2012
|||||
Qy 411 -----IleAspLeuGlyAsnSerHisAlaThrMetLeu 421
|||||
Db 2013 CTATATAAGGACTGATCGCTTCAGGCAATTCAAAATTCGGAATACGGGCTCTGTTCTCGT 2072
|||||
Qy 422 ThrAlaSerTrpLysArgGlnLeuAsnAsnValLeuHisProGluAsnGlyHisTyr 441
|||||
Db 2073 ACAGTTCGTATAAGATAAA---GAGGAAGGGTGTGTTGTCGCAAGCGGTTCCATCGC 2129
|||||
Qy 442 LeuAspGlyLysIleGlyThrLeuGlyThr-----PheLeuSerSerThrAlaLeu 459
|||||
Db 2130 ATCGGCTATAAAATCTTGGCTACGGAAGGACGCGAGGCTATTTAAGAGATGCATCCGTA 2189
|||||
Qy 460 ---IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGly 478
|||||
Db 2190 CCTGTTAAACCGTCGGGAAATCGCG-----GAAGAAGGTACAAATCTCCTT 2237
|||||
Qy 479 ThrPheIleIleArgGlnAlaGlyTyrThrVal----- 490
|||||
Db 2238 GATGTCATTCGAACGGGGGCGGCGAGTTCGTCGTAATACGCTGCACAAAAGGAGCAG 2297
|||||
```

```
Qy 491 ---AlaArgAsp 493
|||||
Db 2298 CCGCAAGGGAC 2309

RESULT 9
US-09-974-300-475
; Sequence 475, Application US/09974300
; Patent NO. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 3213
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-475

Alignment Scores:
Pred. No.: 0.0111 Length: 3213
Score: 112.00 Matches: 119
Percent Similarity: 35.11% Conservative: 72
Best Local Similarity: 21.88% Mismatches: 187
Query Match: 3.53% Indels: 166
DB: 10 Gaps: 29

US-09-857-669-2 (1-609) x US-09-974-300-475 (1-3213)
Qy 20 TyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPhe---AlaLeuPheLys 38
|||||
Db 1714 TATCGGACCGTGCATTCGGTCTGGCGGATTAAAGACCGGCTATGAAGCATATTATTATC 1773
|||||
Qy 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAsp 58
|||||
Db 1774 AACAACTCTGAGACGGTTTCAACA----- 1800
|||||
Qy 59 ThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGln 78
|||||
Db 1801 -----GACTTTCAGCATATCCGACAAAGCTCTACTTTGAG---CCGCTTACTGTTGAAGAT 1851
|||||
Qy 79 GlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsn 98
|||||
Db 1852 GTCATGCATATCATCGATTTCGAGCAG-----CCGAAAGGC 1887
|||||
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLys 118
|||||
Db 1888 GTCGTCGTTTCAG-----TTTGGCGACAGACTGCGATCACTTCC 1929
|||||
Qy 119 AspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly 138
|||||
Db 1930 GAC-----GAGCTCGCAGAAAGGGC 1950
|||||
Qy 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu-----TyrTyr 155
|||||
Db 1951 GTCGCCATTCTCGGAACATCCCTTGAAGATTTGGACCGCGCGGAGGCGGACAAAGTTC 2010
|||||
Qy 156 ArgAsnAlaLeuGluAsnTrpGln-----GlnProValGlySerAspPheAspGlnAsp 173
|||||
Db 2011 GAACAGAGCGCTTGAATGCTGCAGGTACCCAGCGCTTGA----- 2052
|||||
Qy 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaVal-----ThrArgLysGly 190
|||||
Db 2053 -----AAACGGGCGACATCTGTGGAAGCGCGCTTGAATCGCAGCAATATCGGC 2103
|||||
```

Qy 191 TyrProLeuAla-----LysLeuGlyAsnThrArgAlaAlaValAsnPro 205
|||||
Db 2104 TATCGGTGTCGCGCCCATCGTATGTACTCGCGGCGGTGCGATCGGATCGTCTAC 2163
Qy 206 AspThrAlaThr-----ValAspLeuAsnValValValAspSerGlyArgPro 221
|||
Db 2164 CATGAGCAGCAGGCTCTCCATTATATGAAATGCATAAAATCAATCCGCACACCT 2223
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrClnArgTyrProGluGlnIleValSer 241
::: |||
Db 2224 GTCTGTATC---GATAAATTTTACCGGAAAGAAATAGAGTGATGCAATTTCTGAC 2280
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
|||
Db 2281 GGTGAACGGTCGTCATTCGCGGG-----ATTATG 2310
Qy 262 GlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
::: |||
Db 2311 GAGCACATCGAACGGGGGCGTCCACTCGGGGATTTCGATT----- 2352
Qy 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
|||
Db 2353 -----GCCGTTTACCCCGCAATCCCTTTTCAGAGGACATCAAGAAAAGCTT 2400
Qy 302 GluThr---GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
|||
Db 2401 GAACATACACGATTAGCTAGCAAGGGCTTAACATCTCGCGCTCTCAACATTCAG 2460
Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyr 340
::: |||
Db 2461 TTTGCTCTTCAAAAGATGAGTTTACGCTCTTCAAGTC----- 2499
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
|||||
Db 2500 -----AACCGAGATCGAGCGCAACGGTGCCTTCTC 2532
Qy 361 SerAsnValSer-----TyrAsnArgSerThr----- 369
|||
Db 2533 AGCAAAATCACCGCATTCGATCGGCANATCTCCACAAAGGTGATCTCGGAGGGAAG 2592
Qy 370 -----ThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrp 383
|||
Db 2593 CTTGCGGATATGGCTATACAGAGGCTCCAAAAGAG-----CAGGAAGGCGTA--- 2643
Qy 384 TyrValArg-----AspArgAlaGlyIleAspAlaArgLeuGly 396
|||||
Db 2644 TATGTGAAGTCCCGGTCTTCTCGTTCGGAAGCTTGAAGAGTGGATATCACGCTCGGA 2703
Qy 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp----- 410
|||
Db 2704 CCGGAATGAATCGACCGGG---GAAGTCATGGGCAAGACACACATTCGGAAGGCG 2760
Qy 411 -----IleAspLeuGlyAsnSerHisAlaThrMetLeu 421
|||
Db 2761 CTATATAAGGACTGATCGCTTCAGGCATTCAAATTCGCAACTACGGCTCTGTTCTCTG 2820
Qy 422 ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyr 441
|||
Db 2821 ACAGTTGCTGATAAGATAA---GAGGAAGGCTGTGTATCGCAAGCGGTTCCATCGC 2877
Qy 442 LeuAspGlyLysIleGlyThrThrLeuGlyThr-----PheLeuSerSerThrAlaLeu 459
::: |||
Db 2878 ATCGGTATAAATCTTGGCTCGGAAGGACCGGCGCTATTTAAGAGATGCATCCGTA 2937
Qy 460 ---IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGly 478
::: |||
Db 2938 CCGTGAACCGCTCGGGAATCCGCG-----GAAGAAGGTACAAATCTCCTT 2985
Qy 479 ThrPheIleLeuArgGlyGlnAlaGlyTyrThrVal----- 490
|||
Db 2986 GATGTCATTCGAACCGGGGCGCGAGTTCGTCGTCATACGCTGACAAAAGGGAAGCAG 3045

Qy 491 ---AlaArgAsp 493
|||||
Db 3046 CCGCAAGGGAC 3057
RESULT 10
US-09-332-226-1
; Sequence 1, Application US/09332226
; Patent No. US20020025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: InClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/363,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3286 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhea
; STRAIN: FA19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 406..3150
; US-09-332-226-1
Alignment Scores:
Pred. No.: 0.0115 Length: 3286
Score: 112.00 Matches: 150
Percent Similarity: 34.09% Conservative: 89
Best Local Similarity: 21.40% Mismatches: 281
Query Match: 3.53% Indels: 181
DB: 10 Gaps: 36
US-09-857-669-2 (11-609) x US-09-332-226-1 (1-3286)

US-09-871-212-1

```
: Sequence 1, Application US/09871212
: Patent No. US2002034519A1
: GENERAL INFORMATION:
: APPLICANT: Tikoo, Suresh
: APPLICANT: Bablu, Lorne
: APPLICANT: Zhang, Linong
: APPLICANT: Wu, Qiaohua
: TITLE OF INVENTION: MODIFIED BOVINE ADENOVIRUS HAVING
: FILE REFERENCE: 293102003000
: CURRENT APPLICATION NUMBER: US/09/871.212
: CURRENT FILING DATE: 2001-05-31
: PRIOR APPLICATION NUMBER: 60/208,678
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 34446
: TYPE: DNA
: ORGANISM: Bovine Adenovirus 3
: US-09-871-212-1

Alignment Scores:
Pred. No.: 0.557 Length: 34446
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 10 Gaps: 36

US-09-857-669-2 (1-609) x US-09-871-212-1 (1-34446)

QY 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28163 ATTAACCCCGGGGGCTCTCGAGACAAACGACCTCGCGTCTCCACCCGCTG 28222
QY 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28223 GCCTCCGATGAGCGCGCAACGTCACGCTCAACATGCTGACGGCTATATACTAAGGAC 28282
QY 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28283 AACAAAGCTA-----GCTGTCAAAGTAGTCCCGGGCTGCTCCCTCGACTCCAAT 28330
QY 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28331 AATGCTCTCCAGGTCCACACAGCGCGGCTACGGTAACCGATGACAGGTGCTCTCTA 28390
QY 69 GluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThr 88
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28391 AATACCAAGCTCCCTCTCGACACCAACGCGGGCTCTCCCTACTTCTGGTCCCGCAGC 28450
QY 89 GlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28451 CTCCACTAGGTAGGAGGAACAGCTAACACTAACACACC----- 28489
QY 109 PheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28490 -----GGAGCGGGCTCCAAATAGCAATTAACGCTCGGCGGTAAAGTAGGTTC 28540
QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28541 GGTATCACCGTAGTGTCAAAACACAGCTCGCTGCATCCCTGGGGAGCGGTCTA----- 28594
QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28594 ----- 28594
QY 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----Ala 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28595 -----GAAAGCAGAGATAATAAACTGCTGTTAAGGTGGGCGCGGACTTACA 28642

QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 28643 ATAATAATCAAGCTCTTACTGTTGCT---ACCGGGAAC---GGCCTTCAGTCAACCCG 28696
QY 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 28697 GAA---GGGCAACTGCAGCTAAACATTACTCCGCTCAGGCGCTCAACTTTGCAAAACA 28753
QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 28754 AGCCTCCCGG---GAGCTGGGCTCGGCGCTG---CAT 28786
QY 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu----- 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28787 TTTCCCTCCGCGCAAAACCAAGTAAGCTTTATCCCGGAGATGGAATAGACATCCGAGAT 28846
QY 257 -----LeuLeuAspPheGlnGlnAlaLeu 264
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 28847 AATAGGTGACTGTGCCCGCTGGCCAGGCTGAGAATGCTCAACCACTTGCCTGA 28906
QY 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28907 GCT-----TCCGAGACGGTTTGAAGTCCACGACGACCCCTC----- 28945
QY 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
   ||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 28946 --CGGTTAAAGCTCTCCACGCGCTGACATTTGAAATGCGCGCTACGAGCAAACTA 29002
QY 302 GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29003 GGACGAGACTTGGCAGACGAGCTCTGCTCGTCCGTTGTCGACAGGTTCGAGGACTT 29062
QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal 333
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29063 AGATTGCAACGCGCAAGTCCAGATCTTCAGCGGAAGGACCGCCTCGGACT--- 29119
QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29120 -----GATAGCAGCTCTCACTCTCAACATCCGCGCGCCCTACAA 29158
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu 373
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29159 TTTTCTGGACCGCTCTGACTGCTAGTTTGAAGGACGAGTGTGCGATTACTTACAACAGC 29218
QY 374 GluLysArgAlaPhe-----SerGlyGlyIleTyrTyrValArgAspArgAla 389
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29219 AACAAATGGCACTTTCGGTCTCTATAGGCCCCGAATGTGGTAGACCAAAACAGACTT 29278
QY 390 GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAla 401
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29279 CAGGTAAACCCAGCGCTGGTGTAGTCTTCCCAAGMAAACACCTTGTCCCAACCTTGGC 29338
QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer 416
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29339 GATCCGCTGGCTATTTCGACAGCAAAATAGTCTCAGTCTCGGTCGCGCTGACCCAA 29398
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValIleHisPro 436
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29399 GCTTCCCAACGCGCTTAACTTAACT-----TTAGGAACCGGCTTGAATTC 29443
QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29444 TCCAAT---CAAGCGTGTGCTATAAAGCGGCGGGGCTTA---CGCTTTGAGTCTTCC 29497
QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29498 TCACAAGCTTTAGAGAGCAGCCTCACAGTCGGAATAGGCTTAAAGC----- 29542
QY 477 LeuGlyThrPheIleArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn--- 494
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29543 CTTACCGATAGTGTGATCCGCCCAACCTAGGCGACGCGCTCAGAGGTGAGAGCAATAAA 29602
QY 495 -----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArg 512
```

```
Db 29603 ATCATTTGTTAAGCTGGCGGCAATCTTCGTTTAAAAACGGAGCGGTAAACGCC----- 29656
Qy 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
Db 29657 -----GGACCCGTTAACCTTCGCGCCGAGGACCACT 29695
Qy 531 -----GluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
Db 29696 CTCACCTGCAGAACCCCTCGAGCCCTCCAACCTCCCATCTTCAACTGCTCCCTATCG--- 29752
Qy 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
Db 29753 -----GAGGGCTGGTGTGCATAACCAACGCCCTTCTCCAACTGGGAGCGCATG 29806
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArg 578
Db 29807 GAAGTAATACACCGGACTTACTTTAAGACTAGCTCGGTTTGCAAAATCGGT 29860

RESULT 12
us-09-864-761-4449/c
; Sequence 4449, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4449
; LENGTH: 1969
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005366.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
us-09-864-761-4449

Alignment Scores:
Pred. No.: 0.013 Length: 1969
Score: 108.50 Matches: 135
Percent Similarity: 30.84% Conservative: 75
Best Local Similarity: 19.82% Mismatches: 184
Query Match: 3.42% Indels: 287
DB: 10 Gaps: 35

US-09-857-669-2 (1-609) x US-09-864-761-4449 (1-1969)
Qy 25 AspLeuSerGluAsnLysAlaAlaGly-----PheAlaLeuPheLysAsnLysSerPro 42
Db 1868 GAGATTACGAGCGCGTGGCTCCGGGAGCGGCTTCCGCTCGAGAGCGCGCAGCATCCC 1809
Qy 43 AspThrGluSerValLysLeuLys-----ProLysPheProVal 55
Db 1808 GATGTGGGAACAACCTCTTTACAAACCTATGAGCTGAGCGGAAATGAATACTTTGCGCTT 1749
Qy 56 ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIle 75
Db 1748 CGCTGTCAGACGCGGAGGAGCAGCACCAG-----TACGCGAGCTGTGTG 1704
Qy 76 ThrGlnGlnGlnGluValLeuAspLysGluGln----- 87
Db 1703 TTG-----GAGCGCGCCCTGGACGACGAGGAGGAGCTAGTCTCCAGTTAGTGTG 1653
Qy 88 -----ThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThr 101
Db 1652 ACGCGTTGGAGGAGGAGGAGCCAGCTCTCTCCGCGCAGCTTATTCACATCAAGGTG 1593
Qy 102 Met-----LeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 1592 CTGACGCGGAATGACATGCGCCCTGTCTTCAACCAAGTCTTGTACCGGCGCGGTCTGTG 1533
Qy 115 LeuThrGluLysAspGlyAlaTyrThrValHisIle-----ThrProGly 129
Db 1532 GAGGATGCACCCCTCCGCGACCGCGGTGTACAAAGTCTTGCACGCGATCTGGATGAAGGC 1473
Qy 130 ProArgThrLysIle-----AlaAsnValGlyValAla 140
Db 1472 CCCAACGGTGAATATTATTTACTCTCGGACGACCAACCGCGCGGTGCGGCAACTA 1413
Qy 141 IleLeuGlyAspIleLeuSer-----Asp 148
Db 1412 TTCGCCTTAGACCTTGTAAACCGGGATGCTGACATCAAGGTCGCGTGGAGCTTCGAGGAC 1353
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 1352 ACCAACTCCATGAGATTTACATCCAGGCCAACAGACAGGCGCCATCCCGAGGAGCA 1293
Qy 169 -----AspPheAspGlnAspSerTrpGluAsnSerLys 179
Db 1292 CATTCGCAAGTGTGGTGGAGGTGTGTGATGTGAATGACACGCGCGGAGATCACACTC 1233
Qy 180 ThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199
Db 1232 ACCTCCGTGTACACGCCCACTACCGAGGAT-----GCCCTCTCGGAGCTGTC 1185
```



```
QY 200 ArgAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly 219
    |||  ::  |||
Db 1184 ATCGCTTTGCTC-----AGTGTGACTGACCTG-----GATGCTGC 1149
    |||  ::  |||

QY 220 ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle 239
    |||  ::  |||
Db 1148 GAG----- 1146

QY 240 ValSerGlyLeuAlaArgPheGln-----ProGlyThrProTyrAspLeuAspLeuLeu 257
    |||  ::  |||
Db 1145 ---AACGGCGGTGACCTGCGAAGTTCCACCGGTCTCCCTTTGACGCTTACTCTTCC 1089
    |||  ::  |||

QY 258 LeuAspPheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
    |||  ::  |||
Db 1088 CTCAAG-----AATTACTTCACTTTGAAACCACTGCA 1056

QY 278 AspPheAspArgLeuGlnGlyAspArgValPro----- 288
    |||  |||  |||  |||
Db 1055 GACCTGGATCGG-----GAGACTGTGCCAGAAATACAACCTCAGCATCACCGCCGA 1005

QY 289 -----ValLysValSerValThrGluVal 296
    |||  |||  |||  |||
Db 1004 GACGCGGAACCCCTCCCTCTCAGCCCTTACAATAGTCGGTTCACAGTGTCCGACATC 945

QY 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLys 316
    |||  ::  |||
Db 944 AATGACAACCCCTCCACAATCT-----TCTCAA 918

QY 317 IleAlaTyrAspTyrTyr-----AsnLeuPheAsnLysGlyTyrIleGlySer 332
    |||  |||  |||  |||
Db 917 TCTTCTAGCAGGTTTACATTGAAGAAAACAACCTCCCGGGCTCCAATACTATAACCTA 858

QY 333 ValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg 352
    |||  |||  |||  |||
Db 857 AGTGCTGGGACCCGAC-----GCCCGCGAG 831

QY 353 AsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
    |||  |||  |||  |||
Db 830 AATGCTCGGCTTCTTTCTTCTC----- 807

QY 373 LeuGluLysArgAlaPheSerGly-----GlyIleTrpTyrValArgAspArg---Ala 389
    |||  |||  |||  |||
Db 806 TTGGAGCAAGGAGCTGAACCGGGCTAGTGGCTCGCTATTTCACATAAATCGTGACAAAT 747

QY 390 GlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySer 409
    |||  |||  |||  |||
Db 746 GCATAGTGTATCTTA-----GTGCC----- 723

QY 410 AspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu 429
    |||  |||  |||  |||
Db 722 ---CTAGACTATCAGGATCGCGGGGAATTTGAATTAACAGCT----- 684

QY 430 LeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThr 449
    |||  |||  |||  |||
Db 683 -----CATATCAGCGATCGG----- 669

QY 450 LeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhe 469
    |||  |||  |||  |||
Db 668 -----GGCACCGCGGTCTAGCCCAACCAATCAGCTGACATATATT 627

QY 470 PheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThr 489
    |||  |||  |||  |||
Db 626 GTCACCTGATCGCAAT----- 612

QY 490 ValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAla--- 508
    |||  |||  |||  |||
Db 611 -----GACAATGCC-----CCCCAGTCTATATCTCTCGGCCAGGTGGAGCTCG 567

QY 509 -----SerSerValArgGlyTyr 514
    |||  |||  |||  |||
Db 566 GTGGAGATGCTGCCTCGAGGTACTCTAGTGCCACCTAGTGTGACGGGTGTAGCTGG 507

QY 515 GluLeuAsp-----SerIleGlyLeuAlaGly---ProAsnGly 526
```

```
Db 506 GACGGGATGCAGGGCACAAATGCTGGCTCTCTACAGTCTCTTTGGATCCCTTAACGAG 447
    |||  |||  |||  |||  |||  |||
QY 527 SerValLeu----- 529
    |||  |||  |||  |||  |||  |||

Db 446 AGCCTTTTTCATAGGGGTGCACACTGTGTCAAATCAGTACTGCCGTCCAGTCCAAGAC 387
    |||  |||  |||  |||  |||  |||

QY 530 -----ProGluArgAlaLeuLeuVal-----GlySerLeuGluTyrGlnLeu 543
    |||  |||  |||  |||  |||  |||

Db 386 ACAGATTACCCAGGACAGACTCTCAGCGTCTTGATCAAAGACAAATGGGAGCGCTTCGCTC 327
    |||  |||  |||  |||  |||  |||

QY 544 ProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsn 563
    |||  |||  |||  |||  |||  |||

Db 326 TCCACCACTGCTACCTCACTGTGTAGTAACCGAGGACTCTCTCTGAAGCCGAGCCGAG 267
    |||  |||  |||  |||  |||  |||

QY 564 Phe 564
    |||
Db 266 TTC 264
```

RESULT 13

```
US-09-864-761-21197/c
; Sequence 21197, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21197
; LENGTH: 2004
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005366.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EST HUMAN HIT: BF206078.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P33450, EVALU0 4.00e-61
; OTHER INFORMATION: NT HIT: g111416576, EVALU0 0.00e+00
US-09-864-761-21197

Alignment Scores:
Pred. No.:      0.0134      Length:      2004
Score:          108.50      Matches:     135
Percent Similarity: 30.84%      Conservative: 75
Best Local Similarity: 19.82%      Mismatches: 184
Query Match:      3.42%      Indels:    287
DB:              10      Gaps:       35

US-09-857-669-2 (1-609) x US-09-864-761-21197 (1-2004)
QY 25 AspLeuSerGluAsnLysAlaAlaGly-----PheAlaLeuPheLysAsnLysSerPro 42
DB 1764 GAGATTAGGAGCGCGCTGCTCGGGGAGCGGCTTTCCGCTCGAGCGCGCAGCATCCC 1705
QY 43 AspThrGluSerValLysLysLys-----ProLysPheProVal 55
DB 1704 GATGCGGAGCAACTCTTTACAACTATGAGCTGAGCGCGGAATGAATACTTTGCGCTT 1645
QY 56 ArgIleAspThrGlnAspSerGluLeuLysAspMetValGluGluHisLeuProLeuIle 75
DB 1644 CGCGTCAGACGGCGGAGGAGCAGCACAAG-----TACGCGGAGCTGGTG 1600
QY 76 ThrGlnGlnGlnGluValLeuAspLysGluGln----- 87
DB 1599 TTG-----GAGCGCGCCCTGGACCGAGAACGGAGCGCTAGTCTCCAGTTAGTGTG 1549
QY 88 -----ThrGlyPheLeuAlaGluAlaProAspAsnValLysThr 101
DB 1548 ACGCGCTTGACAGGAGGACCCAGCTCTCTCGCCAGCGCTGCTATTTCATCAAGGTG 1489
QY 102 Met-----LeuArgSerLysGlyTyrPheSerSerLysValSer 114
DB 1488 CTGGACGCGAATGACAATGCGCTGTCTTCAACCACTCCTTGTACCGGCGCGCTCCTG 1429
QY 115 LeuThrGluLysAspGlyAlaTyrThrValHisIle-----ThrProGly 129
DB 1428 GAGGATGCACCTCCCGGCACGCGGTGTACAAAGTCTTTCACACGGATCTGGATGAAGC 1369
QY 130 ProArgThrLysIle-----AlaAsnValGlyValAla 140
DB 1368 CCCAACGGTGAATATTATTACTCTTCGCGAGCCACACCGCGCGGTGCGGCAACTA 1309
QY 141 IleLeuGlyAspIleLeuSer-----Asp 148
DB 1308 TTGCGCTTAGACCTTGTAAAGGGATGCTGACAATCAAGGTGCGGTGACTTCGAGGAC 1249
QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
DB 1248 ACCAACTCCATGATGATTTACATCCAGGCCAAAGACAGCGGCCCAATCCCGAAGGAGCA 1189
QY 169 -----AspPheAspGlnAspSerTrpGluAsnSerLys 179
DB 1188 CATTCGAAAGTGTGGTGGAGGTCTGTGGATCTGAATACACACCGCCCGGAGATCAGATC 1129
QY 180 ThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199

```

```

DB 1128 ACCTCGCTGTACAGCCAGCTACCGGAGGAT-----GCGCTCTGGGGAGTGT 1081
QY 200 ArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly 219
DB 1080 ATCGCTTTGCTC-----AGTGTGACTGAGCTG-----GATGCTGGC 1045
QY 220 ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle 239
DB 1044 GAG----- 1042
QY 240 ValSerGlyLeuAlaArgPheGln-----ProGlyThrProTyrAspLeuAspLeuLeu 257
DB 1041 ---AAGCGGCTGTGCTACCTGCGAAGTTCACCGGGTCTCCCTTTACAGCTTACTTCTCC 985
QY 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
DB 984 CTCAG-----AATTACTTCACTTTGAAACCACTGCA 952
QY 278 AspPheAspArgLeuGlnGlyAspArgValPro----- 288
DB 951 GACCTGGATCGG-----GAGACTGTGCCAGATAACAACCTCAGCATCAGCGCCGA 901
QY 289 -----ValLysValSerValThrGluVal 296
DB 900 GACGCGGAACCCCTTCCCTCTCAGCCCTTACAATAGTGTGTTCAAGTGTCCGACATC 841
QY 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLys 316
DB 840 ATGACACACCCCTCCCAATCT-----TCTCAA 814
QY 317 IleAlaTyrAspTyrTyr-----AsnLeuPheAsnLysGlyTyrIleGlySer 332
DB 813 TCTTCTCAGCAGCTTTACATTGAGNAACAACCTCCCGGGGCTCAATACTAAACCTA 754
QY 333 ValValTrpAspMetAspLysTyrGluThrLeuAlaAlaGlyIleSerGlnProArg 352
DB 753 AGTGTCTGGGACCCGAC-----GCCCGCGAC 727
QY 353 AsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
DB 726 AATGCTCGGCTTCTTTCTTCTC----- 703
QY 373 LeuGluLysArgAlaPheSerGly-----GlyIleTrpTyrValArgAspArg---Ala 389
DB 702 TTGGACCAAGGAGCTGAAACCGGCTAGTGGTGTCTATTTCACAATAAATCGTACAA 643
QY 390 GlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySer 409
DB 642 GGCATAGTGTCAATCCTTA-----GTGCC----- 619
QY 410 AspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu 429
DB 618 ---CTAGACTATGAGATCGCGGGAATTTGAATTACAGCT----- 580
QY 430 LeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThr 449
DB 579 -----CATATCAGCGATGG----- 565
QY 450 LeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhe 469
DB 564 -----GGCACCCCGCTCTAGCCACCAACATCAGCGGTGAACATATTT 523
QY 470 PheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThr 489
DB 522 GTCACGTATCGCAAT----- 508
QY 490 ValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAla--- 508
DB 507 -----GACAAATGCC-----CCCCAGGTCTATATCTCGGCGAGGTGGAGCTCG 463
QY 509 -----SerSerValArgGlyTyr 514

```



```
Oy 354 ---TyrArgGlyAsnTyrTrpThrSerAsn----- 362
Db 3448 TTACTTAACCCAAATCATCGGAACAGCAGTCTCTGTAACCTGGATTGGTATTGGTGGTAAA 3507
Oy 363 -----ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPhe----- 378
Db 3508 GGTATGTTTATTAT-----TCAACAGTGGTAAACCAAGCTAAAAAATGCTTCATTAGC 3561
Oy 379 SerGlyGlyIleTyrValArgAspArg-----AlaGlyIleAspAla 393
Db 3562 TTAGGAATAATGGTATTATTTCGATAAATACGGTTATATGGTCACTGGTGTCAATCA 3621
Oy 394 ArgLeuGluAlaGlu-----PheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 3622 ATTAACGGTCTAATATTATTCTTCAATGGTATTCATTAAGAAATGCTATTAT 3681
Oy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 3682 GATAATGGTAATAAAGTATTGTCTTATTATGGAATGATGCCGCTGTTAT----- 3732
Oy 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
Db 3733 -----GAAATGGTACTATCTCTTTGGTCAACATGGCGTTATTTCCAA 3777
Oy 452 ThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr 471
Db 3778 AATGGTATTATGGCTGTCGCTTAACACGCTGTCATGCTGCTTCAATATTTTGATGCT 3837
Oy 472 ProGluAsnLysLysLeuGlyThrPheIleIleArg-----GlyGlnAlaGlyTyrThr 489
Db 3838 TCTGGGTTCACGCTAAGACAGATTTATTACAACTGCTGATGGAAGCTGCGTTAC--- 3894
Oy 490 ValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSer 509
Db 3895 TTTGATAGACACTCAGAAATCAAAATTTCAATCGTTTGTAGAAAT-----TCC 3945
Oy 510 SerValArgGlyTyrGluLeuAspSerIleGlyLeuAla----- 522
Db 3946 AAGGAGAAATGGTCTTATTATTGATCAACAATGGTGTGCTGTAAACCGCTACTGTAAACGCTC 4005
Oy 523 -----GlyProAsnGlySerValLeuProGluArgAlaLeuLeu 535
Db 4006 AATGGACAACGCTCTTTACTTTAAACCTTAATGGT-----GTTCAAGCCAAAGGAGAAATTT 4059
Oy 536 Val-----GlySerLeuGluTyrGlnLeuPro----- 544
Db 4060 ATCAGAGATGCAATGGATATCTAAGATATATTATGATCCTAATTCGGGAATGAGTTTCGT 4119
Oy 545 -----PheThrArgThrSerGlyAlaValPhe---HisAspMetGlyAspAlaAla 561
Db 4120 AATCGCTTTGTTAGAAATTCCAAGGAGAAATGGTCTTTATTGATCACAATGGTATCGCT 4179
Oy 562 AlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 4180 GTAACGTGCCAGAGTTGTTAATGGACAGCGCTCTATTTTAAAGTCTAATGTTGTTTC 4236
```

RESULT 15

US-09-815-242-8513

; Sequence 8513, Application US/09815242

; Patent No. US2002061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

```
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8513
; LENGTH: 9477
```

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(9477)

; US-09-815-242-8513

Alignment Scores:

Pred. No.:	0.147	Length:	9477
Score:	108.50	Matches:	91
Percent Similarity:	34.81%	Conservative:	66
Best Local Similarity:	20.18%	Mismatches:	159
Query Match:	3.42%	Indels:	135
DB:	10	Gaps:	20

US-09-857-669-2 (1-609) x US-09-815-242-8513 (1-9477)

```
Oy 27 SerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGluSer 46
Db 3010 TCITTAATACAGCAATGGGTAACTTAATP-----ANTGGATTGGAGATCATCAAGCC 3063
Oy 47 ValLysLeuLysProLysPheProValArgIleAspThr---GlnAspSerGluIleLys 65
Db 3064 GTTGAACAAGCTGGTAACTTC---ATCAATGCTGATCTATAAACAACAACTGCTTACT 3120
Oy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGln----- 79
Db 3121 ACAGCGTAATAAGACAGAGAGCCATGATTAAACAACAACAACTGCTCAAAATGCGAACCAA 3180
Oy 80 -----GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaPro 96
Db 3181 CCAGAACTAGACAGCAAGCTATTACTAAAGTTCAAACTACACTTCAAGCGTTAAATGGAGAT 3240
Oy 97 AspAsnValLys----- 103
Db 3241 CATAACTTACAAGTTGCTTAAACAACAAATGCGACGCAAGCAATGATGCTTTAAACAAGCTTA 3300
Oy 104 ArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr 123
Db 3301 AATGATCCTCAAAAACAGCATTTAAAGACCAAGCTTACTGCTGCAACTTTAGTAACGTCA 3360
Oy 124 ValHis---IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 3361 GTTCATCAATTTGACCAAAATGCGCAATACG-----CTTAACCAAGCAATGCGTGTAA 3414
Oy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 3415 AGAGAAAGTATTCAAGAT---AACGCAGCAACTAAAGCAAAATAGCAAAATATATCAACGAA 3471
Oy 163 GlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerVal 182
Db 3472 GATCAACCAAGACAAACAAACTAGCATCA-----GCAGTG 3507
```

Qy	183	LeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAla	202
Db	3508	CAAGCTGCA	3525
Qy	203	ValAsnProAspThrAlaThrValAspLeuAsnValVal	216
Db	3526	ATCAATGACGCAACACGACATAGTAAATAATGCGATTAAATCAAGCAGCGCAACTGTG	3585
Qy	217	AspSerGlyArgProIleAlaPheGlyAspPheGlyThrGlnArgTyrPro	236
Db	3586	AATACACGAAAGCAGCATTACATGCTGATGTGAAGTTACAAATGATAAAGATCATGCT	3645
Qy	237	GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu	256
Db	3646	AAACAACGGTTAGTCANATGACATATCTAAACAATGCACAAAACATATATGGAAGATACG	3705
Qy	257	LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln	276
Db	3706	TTGATTGATAGTGAACAACACT	3726
Qy	277	AlaAspAspArgLeuGlnGlyArgValProValLysValSerValThrGluVal	296
Db	3727	-----AGCATAGCAGTTAAGCAAGATTTCGACTGAAGCA	3759
Qy	297	Lys-----ArgHisLysLeuGluThrGlyIle	305
Db	3760	CAAGCATTAGATCAACTTATGAATCATTTACAAAGATTTGCTGCAAAAGATGCAACA	3819
Qy	306	ArgLeuAspSerGluTyr-----GlyLeuGlyLysIleAlaTyrAspTyr	322
Db	3820	CGTCCGAGCAGTGCATATGTCATCGAAGCCGATAAANAACACGATATGAT	3873
Qy	323	AsnLeuPheAsnLysGlyTyrIleGlySerValValThrAspMetAspLysTyrGluThr	342
Db	3874	-----GAAGCAGTTCAAAATGCTGAG	3894
Qy	343	ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTriThrSerAsn	362
Db	3895	TCATCATTCGACGATTAATAATCCGACTATCATTAAGGAAT-----GTATCAAGT	3948
Qy	363	ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIle	382
Db	3949	CGCACACAGCAGTGAACACATCTAAAT	3978
Qy	383	TriPTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu	402
Db	3979	-----GGTTTAGAT	4008
Qy	403	GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr	422
Db	4009	GATAGGCAACACCTGGAATTTCTCTG-----ATCATTTAGATCAATTAACA	4056
Qy	423	AlaSerTriPylsArgGlnLeuLeuAsnVal	433
Db	4057	CCAGCTCAACAACGCGTTTGAAGAAATCAAAAT	4089

RESULT 16

```

US-10-044-090-98
; Sequence 98, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 4065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;

```

```
Qy 277 laAspPheAspArgLeuGlnGlyAspArgValPro-----ValysValSerValt 294
Db 2687 -----AAGCTGCCGATATCATGATCAATCCCAAGACCAATTGCAACGTAGCTGTCA 2735
Qy 294 hrGluVal-----LysArgHisLysLeuGluThrGlyIleArgLeuA 308
Db 2736 TCAACGTGGGGCACCCCGCTGGGATGAACGACGCGCTACGCTGCTGGCGGTGG 2795
Qy 308 spSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLysG 328
Db 2796 GCATTGCCGACGGCCACAGGATGCTGCCATCATATGATGGCTTTGACGGCTTTGCCAAGG 2855
Qy 328 lyTrilleGlySerValValTrp---AspMetAspLysTyrGluThrThrLeuAlaLaG 347
Db 2856 GCCAGATCAAAAGAAATCGCTGGACAGATGTCGGGGCTGCACCGCCCAAGGAGCTCCA 2915
Qy 347 lyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnA 367
Db 2916 TTCCTGGGACAAACCGCTTCTCCCGGGGAAGTACTTGGAAAGAGATCCCCACACAGATGC 2975
Qy 367 rgSerThrThrGlnAsnLeuLysArgAlaPheSerGlyGlyIleTrpTyrValArgA 387
Db 2976 GC-----ACCCACGACATCAACCGCTGCTGATCATCGGT----- 3010
Qy 387 spArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLys---- 405
Db 3011 -----GGATTGAGGCCCTACCTGGGACTCCTGGAGCTGTCAGCGCCCGGAGGAAGC 3062
Qy 406 ----- 428
Db 3063 ACGAGGAGTTCGTGTCCTCCCATGTCATGTTCCCGCTACTGTGTCCAACAATGTGCCGG 3122
Qy 408 lySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArg 428
Db 3123 GTTCGGATTTCAGCATCGGGGCANACACCGCCCTCAACACTATCACCCACACCTGGCACC 3182
Qy 428 lnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyT 448
Db 3183 GCATCAAGCACTCCGCCAGCGGAACCAACGCGCGTTCATCATCGAGACCATGGCGG 3242
Qy 448 hrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyT 468
Db 3243 GCTACTGTGGC---TACTGTGCCAACATGGGGGGCTGCGGGCTGGAGCTGATCCCGCAT 3299
Qy 468 yrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg-----GlyG 485
Db 3300 ACATTTTCGAGAG-----CCCTTCGACATCAGGATCGCAGTCCA 3341
Qy 485 lnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 505
Db 3342 ACGTGGAGCACCTCAGCGGAGAAAATGAAGACCACCATCCAGAGAGGCTTGTGCTCAGAA 3401
Qy 505 erGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle---GlyLeuAlaGlyP 524
Db 3402 ATGAGAGCTGCAGT-----GAAAACTACACCCAGCTTCATTTACCAGCTGTATTTCAG 3455
Qy 524 roAsnGly---SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr---- 541
Db 3456 AAGAGGGCAAGGGGTGTTTCACCTGCAGGAGAACCTGCTGGGTACATGCAGCAGCGGTG 3515
Qy 542 -----GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 560
Db 3516 GGGCACCTCTCCATTTTGATAGAAACTTTGGAACCAAAATC-----TCTGCCA 3563
Qy 560 laAlaAlaAsnPheLysArgMetLysLysLeuLysHisGlySerGlyLeuGlyValArgTrp- 579
Db 3564 GAGCTATGAGTGGATCACTGCACAAACTCAAGAGGCGCCGGGCGAGGAAAAAATTTA 3623
Qy 580 -----PheSerProL 583
Db 3624 CCACCGATGATTCCATTGTGTGCTGGGAATAAGCAAAAGAAAGCTTATTTTCAACCTG 3683
Qy 583 euAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisI 603
```

```
Db 3684 TGGCAGAGCTGAAGAAGCAAAACGGATTTTGACGACAGGATTCCCAAGAACAGTGGTGGC 3743
Qy 603 leSerLeu 605
Db 3744 TCAAGCTA 3751
RESULT 17
US-09-954-456-1601
; Sequence 1601, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1601
; LENGTH: 14800
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1601
Alignment Scores:
Pred. No.: 0.379 Length: 14800
Score: 107.50 Matches: 119
Percent Similarity: 35.17% Conservative: 85
Best Local Similarity: 20.52% Mismatches: 230
Query Match: 3.39% Indels: 147
DB: 10 Gaps: 29
US-09-857-669-2 (1-609) x US-09-954-456-1601 (1-14800)
Qy 64 ileYaspMetValGluGluHis---LeuProLeuIleThrGlnGlnGluVal 82
Db 12286 ATCATCCACCTCGAGAGAGACCCACCGCTGCCCTGGAGTGGCTACAGCGCGGCTC 12345
Qy 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
Db 12346 TTCATGAGGAGGAACAGGAGATCTTCACCGCACCCCTCGGACGAC----- 12390
Qy 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGlyLysAspGlyAlaTyr 122
Db 12391 -----ACCAAGGGCTTCTTTGACCTAAC-----ACGGAGGAGAACCTCACCTAC 12435
Qy 123 -----ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly--- 138
Db 12436 CTGCAGCTGATGGAGCGCTGTGTATCATCACCACCCAGAGCGGCTGTGTCTTCCCGCTG 12495
Qy 138 ----- 138
```

51 proLyspheProValArgatLeAsoThrGlnAsps

```
Db 250 GACCAGTACACTTTGAAAAAGAGTACTCTGCACAAACAGGGTC-----GACCTTGTCATCG 306
Qy 71 HisLeuProLeuIleThrGlnGlnGluValLeuAspLysGluGlnThrGlyPhe 90
Db 307 ATTTTCGACATCATTAAGAAGTCAAAAAGAACCCAGCAAAAAGAGAAAAGACAAG 366
Qy 91 LeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSer 110
Db 367 CAGGCCCATCTGAAAAAGACACAGATCAAAATCTGTT-----GAAGAAAAGCTGACT 417
Qy 111 SerLysVal-----SerLeuThrGluLysAspGlyAlaThrValHisIleThrPro 128
Db 418 TCCGATGTGAGGAGTCTCGCTGTACAGGATTCAGTCAAAACCTGCTCGAGGCCAGCAC 477
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 478 GGAGAGCTTCCATGAGCCAGGATTCGGTCATCACCCCGCTCAATGACGTGATGAGCAAA 537
Qy 149 -----GlyAsnLeuAlaGluTyrTyrArgAsnAla----- 158
Db 538 GAAATCCGCGCAGCAAGATCGATCAAGCAAGAACATGTAACCAAGAACTCAAGGGC 597
Qy 159 -----LeuGluAsnTrpGlnGlnProValGlySer----- 168
Db 598 AATTGATACCTTCAAAATTAATAAAACCCCGCTGATGAAATCGGACGCTTTTTCGATCAT 657
Qy 169 -----AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 658 CCCAATTACGTATACGTCTGACAAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 717
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 718 GTA-----CAGCAGGTACAGATCAACGAC 741
Qy 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 742 GGCAGATACCTGTTGAGGAAATACAGTATCGACAGCGGAA-----GTCTATCGA 792
Qy 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 793 AAGCTCAATTCAGAGG-----CTCTGAGCGGAGCCCAATC 831
Qy 246 PheGlnPro-----GlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
Db 832 TTCAAACCTGTGCGCGGACTG-----CTGCTTTTAATCGCGCTTTTCATCTCAGCT 882
Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 883 TTGGTT-----TACTAT-----TTCGAAAAGCAATAC 909
Qy 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
Db 910 AGCAGCGGATTCGCAAAAACAAGTCG-----CTCATGCTGCTTTTCTGTGATCACA 960
Qy 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsn 323
Db 961 GGGATCTTTTA-----ATCGTCATGGAAGTCATCAGC 993
Qy 324 LeuPheAsnLys-----GlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 994 CTGTTTCAGCAGCTGTAATCGGCCATATCGGC-----TAT 1029
Qy 341 GluThrThrLeuAlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsn 357
Db 1030 CTCGTCGGGATTCGATGGGCGGTGATGCTCATTAATGCTGATCAATGAAGCGCTCGC 1089
Qy 358 TyrTrpThrSer-----AsnValSerTyrAsnArgSerThrThr 370
Db 1090 ATCTGTGCGAGCATGATCTTTCCATTTGCGGAAGCATGATGTTTAAACCAGGGGTGCACA 1149
Qy 371 GlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGly 390
Db -----
```

```
Db 1150 GGCAGTTCAAC-----TATGGAATCGGAACGATATTCTTTA----- 1185
Qy 391 IleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp 410
Db 1186 ATCAGCTCGATGGCGGTATACTGTTTATAGGGAAG----- 1221
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1222 -----CATACCGCAGGCGCAAAATCTCCAGCGCGGTATTTGTCCTCTCATC 1272
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrLeu 450
Db 1273 AACATCTGTTGTT----- 1284
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1285 -----CTGTTTACCTTCACGCTGATTCAAAACAGCGGCAAAACCGGCTT----- 1329
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla-----Gly 487
Db 1330 -----GAGATGGGTACATATCTTGTCTATGGGGCGGTGTCGCGCATCGGA 1374
Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet-----PheArgSer 505
Db 1375 TCGTCAGTT-----CTTGTCTTCGAGCTCATCGCGT-----TTTGAAGC 1416
Qy 506 Gly---GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
Db 1417 GGATTTGGCATCTGTCACG-----ATGAAGCTTCTCGAGCTTTTCAATCCG 1464
Qy 525 AsnGlySerValLeuProGluArgAlaLeuLeuVal-----GlySerLeuGluTyr 541
Db 1465 AATCATCGCTGCTC-----CGCAAAATCTGACGGAACACCGCGGCACATCATCAC 1518
Qy 542 GlnLeuProPheThrArgThrLeuSer----- 550
Db 1519 AGCGTT---ATGPTCGCGAACTTATCTGAGCGGCATCGGAAGCGGTGCGGAACAGGG 1575
Qy 551 -----GlyAlaValPheHisAspMetGly 558
Db 1576 CTGCTGCAAGGTGCGCGGCTATTACCATGATATCGGC 1614

RESULT 19
US-09-974-300-2620
; Sequence 2620, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680, 598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279, 526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2620
; LENGTH: 2140
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2620

Alignment Scores:
Pred. No.: 0.025 Length: 2140
Score: 106.50 Matches: 125
Percent Similarity: 36.02% Conservative: 103
Best Local Similarity: 19.75% Mismatches: 200
Query Match: 3.36% Indels: 205
DB: 10 Gaps: 38
```


COUNTRY: USA
 ZIP: 10014
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/332.226
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/363.124
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/973.336
 FILING DATE: 05-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/572.187
 FILING DATE: 23-AUG-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gallagher, Thomas C.
 REGISTRATION NUMBER: 37,066
 REFERENCE/DOCKET NUMBER: SPA-1-PDC
 TELEPHONE: (212) 645-1405
 TELEFAX: (212) 645-2054
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3537 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORGANISM: Neisseria meningitidis
 STRAIN: FAM18, FAM20, B16B6, group X and group W135
 FEATURE:

NAME/KEY: CDS
 LOCATION: 721..3450
 FEATURE:
 NAME/KEY: mat-peptide
 LOCATION: 793..3447
 US-09-332-226-3

Alignment Scores:
 Pred. NO.: 0.0542 Length: 3537
 Score: 106.50 Matches: 151
 Percent Similarity: 35.03% Conservative: 90
 Best Local Similarity: 21.95% Mismatches: 285
 Query Match: 3.36% Indels: 162
 DB: 10 Gaps: 36

US-09-857-669-2 (1-609) x US-09-332-226-3 (1-3537)

Qy 6 ThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyAlaProAlaAlaAsp 25
 Db ACAGCAACATTGTTCCTGCTAAATATATTTATGCTGCTTTAATAGCCGCGCTG 785
 Qy 26 Leu-SerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGl 45
 Db TTATGCACAAATGTCAAGCGCAACAGCA- ----CAGGAAACACAGTTGGATACCAT 839
 Qy 45 uSerLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLy 65
 Db ACAGGTAAGCCAAA- ----AAACAGAAAAACCGCGCGGATAACGAAGTAACCGGCTGGG 896
 Qy 65 sAspMetValGluGluHisLeuProLeuIleThrClnGlnGlnGluValLeuAsp- 84
 Db CAAGTTGGTCAAGTCTCCGATACGCTAAGTAAAGAACACAGGTTTGAATATATCCGAGACCT 956

Qy 85 ----LysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLe 103
 Db GACCGGTTATGATCGGGTATTGCGGTGGTCAACAGGGTGGGGCGGAAGTTCCGGGTA 1016
 Qy 103 uArgSerLysGlyTyPheSerSerLysValSerLeuThrGluLysAspGly- - - - - 120
 Db TTCAATACGGCGCATGGATAAAACCGGTTTCCTTAACG- - -GTAGACGGCGTTCGCA 1073
 Qy 121 ----AlaTyThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGl 138
 Db AATACAGTCTACACCGCGCGGCGCATTTGGGT- - -GGGACGAGGCGCGGTAGCAG 1130
 Qy 138 yValAlaIle- - - - -LeuGlyAspIleLeuSerAspGl 149
 Db CGCGCAATCAATGAATCGAGTATGAACAGTCAAGCCGTTGAAATCAGCAAGGGTTC 1190
 Qy 149 yAsnLeuAlaGluTyTyArgAsnAlaLeuGluAsn- - - - -TrpGlnGlnProVa 166
 Db CAATTTCATCAGATACGGAACGCGCATTTGGCAGGTTCGCGCATTTCAAAACCAAAAC 1250
 Qy 166 lGlySerAspPhe- - - - -AspGlnAspSerTrpGlu- - -AsnSerLysThrSerValLe 183
 Db CGCAGCGCATTTATCGGAGAGGGAACACAGTGGGGCATTCAGAGTAAACTGCCTATTC 1310
 Qy 183 uGly- - - - -AlaValThrArgLysGlyTyProLeuAlaLysLeuGlyAsnThrAr 200
 Db GGGAAAGACCATGCCCTGACGCAATCC- - - - -CTTGGCTTGGCGGACGCGGG 1361
 Qy 200 gAlaAla- - - - -ValasnProAspThrAlaThrValAspLeuAsnValValVa 216
 Db CGCGCGGAAGCCCTCTTATTATTAACGCGGGGTCGGGAAATCCATTCGCATAA 1421
 Qy 216 lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyPr 236
 Db AGATGCGCGGCAAG- - - - -GGTGTGCAGAGCTTC- - -1449
 Qy 236 oGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyAspLeuAspLe 256
 Db AACCGCGTGTGTGGACGAGGACAGAGGGTGGCAGTCAGTCAGATATTTTCATT 1508
 Qy 256 uLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTySerGlyAlaSerValGl 276
 Db G- - - - -TGCAGAAAGAAATGCCACAATGCATATGCGGCGCTGAA 1547
 Qy 276 nAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVa 296
 Db A- - - - -AACAGCTGAAGAGAGAT- - -GCCTCGGTCAAGATGAGCGCAAAACCGT 1595
 Qy 296 lLysArgHisLysLeuGluThrGlyIleArgLeu- - - - -AspSerGlyTyArgLysLe 313
 Db CAGCAGCAGCATTTATACCGCGCTCCACCCTTACTTCCGAACCCGCTTGAGTAGCGAG 1655
 Qy 313 uGlyGlyLysIleAlaTyAspTyTyAsnLeuPheAsnLysGlyTyIleGlySerVa 333
 Db CCAATCATGGCTGTTCGACCGGGTTGGCATTTGGACAACCGCCATATGTGCGAGCGGT 1715
 Qy 333 lVal- - - - -TrpAspMetAspLysTyGluThrLeuAlaAlaGlyIleSe 349
 Db TCTCGAACGTACGACGACGACCTTTGTATACAGGGGATGACTGTCTGCTATTTTAC 1775
 Qy 349 rGln- - - - -ProArgAsnTyArg- - - - -GlyAsnTyTrpThrSerAsp- - 362
 Db CAGTCAAGATTATGATCCCGGTTCGCTCAAGAGTCTTGGCAAAATATCGGGCGGATAATA 1835
 Qy 363 - - - - -ValSerTyAsnArgSerThrThrGlnAsnLeuGluLysArgAl 377
 Db GGCAGAAAGCGTGTGTGTTTCAGGGAGGAGGAGGAGTACATTCAGAGGTATC- - - - -GG 1886
 Qy 377 aPheSerGlyGlyIleTrpTyValArgAspArgAlaGlyIleAspAlaArgLeuGlyAl 397
 Db TTACGGTACCGCGGCTGTTTAT- - - - -GATGAACGCCATACTATAAAACCCGCTACGGGGT 1940

```
Qy 397 aGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHi 417
|||:||||:
Db 1941 CGAATATGTTTACCAT-----AATGCTGATAAGGATACCTGGGCCGATTA 1985

Qy 417 sAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGl 437
|||:||||:
Db 1986 CGCCCGACTT-----TCTTATGACCGCAA----- 2010

Qy 437 uAsnGlyHisTrpLeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerTh 457
|||:||||:
Db 2011 ---GGTATAGATTGGCAACCGTTTGCACACACGATGCTCTCAGCAGGTTTCGGA 2066

Qy 457 rAlaLeulleArgThrSerAlaArgAlaGlyTrpPhePheThrProGluAsnLysLysLe 477
|||:||||:
Db 2067 TAAAAATTTGCCGTCGCCGACGCAATAAACCGTATTCTTTATATAATCCGACCGGATGAT 2126

Qy 477 u-----GlyThrPheIleIleArgGlyGlnAlaGlyThrValAlaAar 492
|||:||||:
Db 2127 TTATGAAGAAGCGAAGCTGTT-----CAAGCAGTATTTAAAAAGGCATT 2174

Qy 492 gAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly----- 507
|||:||||:
Db 2175 TGATACGGCCCAAAATCGTCAATTTGAGTATCAATCTAGGGTACGACCCCTTTAAGTC 2234

Qy 508 -----AlaSerValArgGlyTrpGluLeu----- 516
|||:||||:
Db 2235 GCAATTGTCCACACAGCGATTATTCTTCAAAAGCGCAGTTCAGGCAATGATTGATAAC 2294

Qy 517 -----AspSerIleGl 520
|||:||||:
Db 2295 CCCGAAAAGCCTCCGTTTCCCAACGGAAGCAAAAGCAACCCGATAGGGTGTCTATCGG 2354

Qy 520 yLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGl 540
|||:||||:
Db 2355 CAAGACACCGTCAATACATCCCGATACCTGGT-----TTCCGCAATAACAC 2402

Qy 540 uTrpGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAl 560
|||:||||:
Db 2403 CTATACAGCTGCACACCGAGGAATATCGCGGCAACGGTTATTAT----- 2448

Qy 560 aAlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPh 580
|||:||||:
Db 2449 -GCAGCCGTTCAAGACAAATGTCCTGTTGGGC-----AGGTGGGC 2486

Qy 580 eSerProLeuAlaProPheSerPheAspIleAlaTrpGlyHisSerAspLysLysIleAr 600
|||:||||:
Db 2487 GGATGTCGGACGACGATACCTTACGATTACCGCAGCAGCATTTCGGAGATAAG----- 2541

Qy 600 gTrpHisIleSerLeuGlyThr 607
|||:||||:
Db 2542 ----AGTGTCTCTACCGCACT 2559
```

RESULT 21

```
US-09-912-020-139
; Sequence 139, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allynn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
```

```
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: E. Coli
US-09-912-020-139
```

Alignment Scores:

```
Pred. No.: 0.0162 Length: 1149
Score: 104.50 Matches: 79
Percent Similarity: 35.31% Conservative: 46
Best Local Similarity: 22.32% Mismatches: 142
Query Match: 3.29% Indels: 87
DB: 10 Gaps: 18
```

US-09-857-669-2 (1-609) x US-09-912-020-139 (1-1149)

```
Qy 232 ThrGlnArgTrpProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrPro 251
|||:||||:
Db 175 AGTTCACGACGACGACCAATATCATTTGGATTA-----AATGTCGCG 219

Qy 252 Tyr---AspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTr 270
|||:||||:
Db 220 TTTGGTGATATTACGACTTCGCTGAATTACAGC-----TAT 255

Qy 271 SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValproValLys 290
|||:||||:
Db 256 TCCAAATATATATGCAAAACGATCGGATCATTTACTCGCTTTTACGCTTAATGTTCC 315

Qy 291 ValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg----- 306
|||:||||:
Db 316 TTCAGT-----CATTGGATGCTACAGACAGTCAGTCGCGCATTTTCGT 357

Qy 307 ---LeuAspSerGluTrpGlyLeuGlyLysIleAlaTrpAspTrpTyrAsnLeuPhe 325
|||:||||:
Db 358 AATTCAAACGCGCTTACAGTATGTCAAACGATTGAAAGGCGCATGACCAATCTA--- 414

Qy 326 AsnLysGlyTrpIleGlySerValValTrpAspMetAsp---LysTrpGluThrThrLeu 344
|||:||||:
Db 415 ---TCGGGGGTTTATGCGACTCTGCTCGCGGATAATAACCTGAATTATAGCGTTTCA 471

Qy 345 -----AlaAlaGlyIleSer-----GlnProArgAsnTrp 354
|||:||||:
Db 472 GGTAAACACCCACGAGGTAAATACATGCTCTCGCACCATGGTTACAGTCTCTTAATTAT 531

Qy 355 ArgGlyAsnTrpTrpThrSerAsnValSerTrpAsnArgSerThrThrGlnAsnLeuGlu 374
|||:||||:
Db 532 CGTGAGCTTATGTTAACTACTAATGTCGGTTACAGTCGAGTGGTGACAGCAGCAGATT 591

Qy 375 LysArgAlaPheSerGlyIleTrpTrpValArgAspArgAlaGlyIleAspAlaArg 394
|||:||||:
Db 592 TATTACCGGAATGAGTGGTGGGATTTATGCTCATGTCATGCGATCACCCTGGACAGCG 651

Qy 395 LeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGly 414
|||:||||:
Db 652 CTGGCGGACACAAATGGTCTCGTT-----AAGCTCTCTGGTGTGAT---AATGTCAAA 702

Qy 415 AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
|||:||||:
Db 703 ATAGAGAACCACGACCGGAATTCATACCGACTGGCGT-----GGCTATGCCATA 750

Qy 435 HisProGluAsnGlyHisTrpLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
|||:||||:
Db 751 TTACCATTTGCGACAGAAATATAGAGAAACCGTGTGCTCTTAACGCGGAATTCCTTCA 810

Qy 455 SerSerThrAlaLeuIleArgThr-----SerAla 464
|||:||||:
Db 811 GATATGTTGAACCTGGATGGAACCGTGTCTACTGTCACTCCCACTCAGCGTGTATTGCC 870

Qy 465 ArgAlaGlyTrpPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
|||:||||:
|||:||||:
```

```

Db      871 AGAGCAACATTT-----AATGCACAATAATCGGG-----GGG   900
Qy     485 GlnAlaGlyThrValAlaArgAspAenAlaAspValProSerGlyLeuMetPheArg   504
       ::          |||         ||||    |||        :||
Db     901 AAAGTATTAAATGACGTGAAGTCAGCTAATAAGACGGCTTCATTCCGGTGCAATTGTGCACA   960
Qy     505 SerGlyGlyAlaSerSerValArgGlyTyrrGluLeuAspSerIleGlyLeuAlaGlypro   524
       |||         |||         |||        :||
Db     961 CACGGA-----GAGAATAAAAATGCCAGCATTTGTCGCGCAA   996
Qy     525 AsnGlySerVal-----LeuProGluUatrgAlaLeuLeuValGlySer-----   538
       ||||||   |||         |||||:::|   |||         |||
Db     997 AATGGTCAGGTTTTATCTGACTGGACTTCACAGTCAGGCCAAATTCACAGGTTTTCATGGGGC   1056
Qy     539 -----LeuGluTyrrGlnLeuPro   544
       :::|||||:::|||||
Db     1057 AAAGATAAAAACTCAAACTGATTGTCGAGTAACAAGCTTCCT   1098

RESULT 22
US-09-974-300-2362
; Sequence 2362, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIORITY FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIORITY FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2362
; LENGTH: 2055
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2362

Alignment Scores:
Pred. No.:           0.0452            Length:           2055
Score:              104.00             Matches:          119
Percent Similarity: 36.26%             Conservative:     100
Best Local Similarity: 19.70%           Mismatches:      195
Query Match:        3.28%               Indels:          190
DB:                  10                Gaps:           36

US-09-857-669-2 (1-609) x US-09-974-300-2362 (1-2055)

Qy     20 TyrAlaProNlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsn   39
       |||||||::|         ::|         |||||
Db     79 TATCGCGCGCGCAC---GTTGAAGACCACAAAAGGCAACGGCT-----AAT   120
Qy     40 LysSerProAspThrrGluservallysLeuLysProLysPheProvalArqIleAspThr   59
       |||         ::|         :|||:::|   :|||         :||
Db     121 AAAAACGAGGACGGCAGAGAG---GAAGTTCGCCGCCACCATCTGAAAAAAAAAGAGTAGTC   177
Qy     60 GluAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln   79
       |||:::|   ::|         |||::|   |||::|   |||::|   :|||         :||
Db     178 TC TGACACACAGGGTC---GACCTTGTCATCGATTTTCGACATCATTAAGAAGAGTAGTCAA   234
Qy     80 GluCuuValLeuAspLysGluInThrGlyPheLeuAlaGluGluAlaProAspAsnVal   99
       :|||         ::|         |||:::|   |||         :||
Db     235 AAAGAAAGCGCAGAAAAAGAGAAAAAAGACACAGCAGGCCCATCTGAAAAAGACCGATC   294
Qy     100 LysThrWetLeuArgSerLysGlyTyrrPheSerSerLysVal-----SerLeuThrGlu   117
       |||:::|   :|||         :|||:::|   |||         :||
Db     295 AAATTCGT-----GAAGAAAGCTGATCCGATGTGACGGAGCTCTGCTGTACAG   345
Qy     118 LysaspclYALAtyrThrValHisIleThrProGlyProArqrThrLysIleAlaAsnVal   137

```

```
Db 1186 -----CTGTTTACCCTGCAGCTG 1203
QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1204 ATTCAAAACACGGCGCAACCGGCTT-----GAGATGGGTACA 1242
QY 480 PheIleIleArgGlyGlnAla-----GlyTyrThrValAlaAlaArgAspAlaAsp 496
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1243 TATCTTGTCTATGGGGGGGTGTCGGCATCGGATCGTCAGTT-----1284
QY 497 ValProSerClyLeuMet-----PheArgSerGly---GlyAlaSerSerValArgGly 513
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1285 CTTTGTCTCGGACTCATCGCGTTTGTGAAGCGGATTTGGCATCTGTCACG-----1338
QY 514 TyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAla 533
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1339 -----ATGAAGCTTCTCGAGCTTTCNAATCCGAATCATCCGTGCTC-----CGCAAA 1386
QY 534 LeuLeuVal-----GlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSer 550
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1387 ATTCTGACGGAACACCGGGGCACATACCATCACAGCGTT--ATGCTCGGGAAGTTATCT 1443
QY 551 -----GlyAlaValPhe 554
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1444 GAGCGGCATCGAAGCGGTGCGAGCAACGGCTGCTGCGAAGGGTGGCGGCTATTAC 1503
QY 555 HisAspMetGly 558
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db -1504 CATGATATCGGC 1515

RESULT 23
US-09-974-300-2632
; Sequence 2632, Application us/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2632
; LENGTH: 4557
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2632

Alignment Scores:
Pred. No.: 0.176 Length: 4557
Score: 103.50 Matches: 130
Percent Similarity: 31.64% Conservative: 94
Best Local Similarity: 18.36% Mismatches: 222
Query Match: 3.26% Indels: 262
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-974-300-2632 (1-4557)
QY 25 AspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysLysSerProAspThr 44
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 571 GAAGTGAAGAAGATTCGAATGTCAAAAGGCGATGTTTGTCAACACCCAGTTCAAGCGCA 630
QY 45 GluSerValLys-----LeuLysProLysPheProValArgIleAspThrGln 60
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 631 CAAAAATAAAGGCTCGAGACTCAAAAGATTCCTTTTCGGTGAATACGACACATCCCG 690
QY 61 AspSerGluIleLysAspMetValGluHisLeuProLeuIleThrGlnGlnGlu 80
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
```

```
Db 691 GAAGTAACGGTT-----CTGTTTACCCTGCAGCTG 702
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 702 -----1284
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr---GluLysAsp 119
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 703 -----TATTTCAACCAACCAAGCAATAGCAAAACCATTTGATAATCA 741
QY 120 GlyAlaTyrThrValHisIleThrProGlyProArgThr-----132
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 742 GGAACCTTTCGATAAAGCATCAACCCCGCGCATGATGGACAGTCAGCTCAACAAA 801
QY 133 -----LysIleAlaAsnVal-----137
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 802 AAATTGGATCAGGTGAANAATGCCAAGCTCGAGAAGCTTCCAGAGGGGTGACCATCGT 861
QY 138 -----GlyValAla 140
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 862 TCTGTCAAAGTTTTATCAGCTGGATGTCACATAGATGTTTCAGTCACGCCGGGAGATCAA 921
QY 141 IleLeu-----GlyAspIleLeuSerAspGlyAsnLeuAla 152
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 922 GTTCTTTCAGGCTACACCGTAGACGAGATGGAATAATGTCACATTCGACGGTGAGATAGAT 981
QY 153 GluTyrTyrArg-----156
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 982 TCGGCTTACCGCTCATCTATGTACGGACATCGATCGCTCCGCAAAACCGAAGAGGC 1041
QY 157 -----AsnAlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSer 174
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1042 GGAACCGCTGCATTTAGCAACACGCGTTCGCGCGAGACATCTGGAACCGGCTCT 1101
QY 175 TrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----189
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1102 GCAGAGCGCTCTGTACGCGCAATACGGAATAATGATCGAAAAATCGTCACAGGTTAT 1161
QY 190 -----GlyTyrProLeuAlaLysLeu 196
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1162 AAAGCGATTACACAGCGTTCAGCTGCGAGCATTTTATATACTATGAGAGAAAAATC 1221
QY 197 GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 216
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1222 GACGAATCCAAAGCGAGCATAACAGACTCGTTGCTCCGCTGATTTGCATCTTTTCA 1281
QY 217 AspSer-----GlyArg 220
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1282 GATTCTTAAAGGTGATTCCGATTACTTCAATCAGAAGCGCAGTCAGCGGTGGAACG 1341
QY 221 ProfileAlaPheGly---AspPheGluIle-----ThrGlyThrGlnArgTyr 235
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1342 CCTCTAAAGCAAGCAAGGATTTATACCGCTCTCTGATAATGAAGCGGCTTTGAAATCAA 1401
QY 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1402 TTTAAACCAAGTGTACCGGAGCTTACAAAATCAGTATCAACCGGAGGTCAACAGCGGA 1461
QY 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerVal 275
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1462 GTAATCATCGAT-----AAATCAACAACCTATACGAACACTGCAGTG 1503
QY 276 GlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGlu 295
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1504 -----ACCGGAACAGGAGAAATCAAAAGAGCTTTCAGGTACAGCC 1542
QY 296 ValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 1543 GTTCAGGAGAAATTTGATAAAGGCTACTCAAAACATTTGATTACAAG-----1587
QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
```

Db 1588 AAGAAACGGTCGATTCGCGAATCAACAAAAAACAACATACAGATGAACAATTGG 1647
Qy 336 AspMet---AspLysTyrGluThr-----ThrLeuAlaAlaGly----- 347
Db 1648 AAGTGGATCACAATTTCGAAGCGGGCGCTCAGCGTCTCTCGACGGATCATTCGCGCTT 1707
Qy 348 -----ileSerGlnProArgAsnTyrArg----- 355
Db 1708 CAGGATGTACGAATAACAAACGCTTGAAGAAGGCAAGCACTATACATGATAAATAA 1767
Qy 356 -----GlyAsnTyrTrpThrSerAsn----- 362
Db 1768 CCGGATCATAAGGTTTCTCTCGAACTCATCGGAGATTATCGCACGAACAGCAGCCAG 1827
Qy 363 -----ValSerTyrAsnArgSer-----ThrThrGlnAsnLeuGluLys 375
Db 1828 CTTAAGATCAGTATACACAGATATGATGCCGACTTTTCCAAACGAANAATGCAAAAAC 1887
Qy 376 ArgAlaPheSerGlyGlyIleThrPtyrValArgAspArg-AlaGlyIleAspAlaArgLe 395
Db 1888 ACGCAGAATCA-----ACGTGG-----ACTGATCAGACGACACTGNACCCAANAAT 1935
Qy 395 u-GlyAlaGlu---PheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuG 414
Db 1936 AAGGAAGCGAGCGAGTTTACGCGGAATCGCGACATCCCA-----C 1977
Qy 414 LysAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnVal 434
Db 1978 AAGCGCTTCAACACGCTTATACACGCTGTTCACAAAGAAAT----- 2021
Qy 434 euHisProGluAsnGlyHisTyrLeu-AspGlyLysIleGlyThrThrLeuGlyThrPhe 453
Db 2022 --CACGTGAAAATCGCATCAACTATAATGGAGAG----- 2055
Qy 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGlu 473
Db 2056 CCGTCAAAAACCCCTATATTAAGATCTTTAGCAGATGACCAATTT----- 2106
Qy 474 AsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgasp 493
Db 2107 -----GAAAAGGTCGGTCTGCTCAAG-----ACGTATTCAGTTAATAAAGAC 2151
Qy 494 AsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGly 513
Db 2152 GOTTCC-----ATCACAGAAGGAGACATCTGCCCGCCCTCCCAA 2190
Qy 514 TyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAla 533
Db 2191 TATGATGTGAAGAGCGCTGCGCTGCC-----AATAACAAACTTTGACCGTCCATTG 2244
Qy 534 LeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaVal 553
Db 2245 AAAACGGACGATTCGCTCCGCTATCTTATGNAATTTAAACATCGCTAAAGGACAGGTC 2304
Qy 554 Phe---HisAspMetGlyAspAlaAlaAlaAsnPhelLysArgMetLysLeuLysHisGly 572
Db 2305 ATTAGCAGCATCTCTTACACCAATAAGGCCACGCTAC-----CATAT 2346
Qy 573 SerGlyLeuGlyValArg 578
Db 2347 GCCGATATTCAGACGCT 2364

RESULT 24

US-09-070-927A-132
; Sequence 132, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 5480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 132:
US-09-070-927A-132

Alignment Scores:

Pred. No.: 0.233 Length: 5480
Score: 103.50 Matches: 145
Percent Similarity: 30.06% Conservative: 88
Best Local Similarity: 18.71% Mismatches: 233
Query Match: 3.26% Indels: 309
DB: 10 Gaps: 38

US-09-857-669-2 (1-609) x US-09-070-927A-132 (1-5480)

Qy 42 ProAspThrGluSerValLys---LeuLysProLysPheProValArgIleAspThrGln 60
Db 2514 CCAGATGTTGACAGCATCGATGGTTTAAAGTCCAGCTATTTCATTAAGCAAAAGACGACA 2573
Qy 61 Asp-----SerGluIleLysAspMetVal----- 68
Db 2574 AGTAAACCCAGCTCAACGGTGGGAGCGGTACAGAAATCAATGATTATTACGATTG 2633
Qy 69 -----GluGluHisLeuProIleThrGln 77
Db 2634 TTATTTGCTCGGTAGGTCAATTTTCTCTAAGCATATATTAATTAACAGTCAG 2693
Qy 78 GlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAsp 97
Db 2694 TCTGTTGAACAGATGGTTGATAAA-----GTCTTAGAGCTACAGAG 2735
Qy 98 AsnValLys-----ThrMetLeuArgSerLysGlyTyrPheSerSer 111
Db 2736 CGAACAAAATTCAAATTTTGGCCCCCTGTTGTCGTCAGAAAAAGGCCAACATAAAAA 2795
Qy 112 LysValSerLeuThrGluLysAspGly-----AlaTyrThrVal 124
Db 2796 GTGTTGAAATGATCCACAGCTGAAGGTTATGTCAGAAATGCGGGTGGATGGAACCTAT 2855
Qy 125 HisIleThrProGlyPro-----ArgThrLysIleAlaAsnValGlyValAlaIle 141


```
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-28

Alignment Scores:
Pred. No.: 0.0245 Length: 1164
Score: 103.00 Matches: 68
Percent Similarity: 36.31% Conservative: 50
Best Local Similarity: 20.92% Mismatches: 114
Query Match: 3.25% Indels: 94
DB: 10 Gaps: 13

US-09-857-669-2 (1-609) x US-09-974-300-28 (1-1164)
Qy 91 LeuAlaGluGluAlaProAspAsnValLysThrMetLeu----- 103
Db 64 CTGGCAGAGAGGTGAACAAATACGGCCCAATGCTCTCGTATACGGCGGAGGCACG 123
Qy 104 ---ArgSerLysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
Db 124 ATTAAGCGAAACGGTTTATACGATCAAGTCATTTCAATCCTTGAAAGCGGCGCG--- 180
Qy 123 ThrValHis-----lleThrProGlyProArgThrLysIleAlaAsnValGly 138
Db 181 ACCGTCCTAAGTACCCCGGCTCGAACCGAATCCGCGTGTGGCCACTGTGAATAAGGA 240
Qy 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAla 158
Db 241 GTTGGATCTGCAAGAGAACCATATTGACTTCTTTGGCAGTCGGCGCGGAGCGTC 300
Qy 159 LeuGluAsnTrpGlnGln---ProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
Db 301 ATTGATTGTACGAAAGCAATTGCTCCGAGCGAATAACGACGGCGATGCTGGGAT--- 357
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro----- 192
Db 358 -----ATTGTGAGAAAACCAATATTTCCGGCTGATCGCGTGGCG 396
Qy 193 -----LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr 209
Db 397 TTTGGAACAGTTTAAACGTTAGCAGCAACAGCGCTCTGAAATCAACTCGGGATCTGTGATC 456
Qy 210 Val-----AspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 457 ACAAAATTGGGAAACCAATGAAAAATACGCGCTGGGGAAGCCGCTGATTCCTAAATTT 516
Qy 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 517 TCATTT-----CTTGA-TCCGGTCAACACGTTTACCGT 548
Qy 248 ProGlyThrProTyrAspLeu-AspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAs 267
Db 549 CCGGAAGACCAACACGATTTCAGGCATGTCGACATGATGCCACGTCCTGTTTGAGCAATA 608
Qy 267 nGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlnArgVa 287
Db 609 TTTTCACCATACC-----GAAATATAC 629
Qy 287 lProValLysValSerValThrGlu---ValLysArgHisLysLeuGluThrGlyIleAr 306
Db 630 CCTTATACGACCGGATGCGCAATCCCTGCTTAAACGGTAATTTGAACAGCTCCTTAA 689
Qy 306 gLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAs 326
Db 690 GCTC----- 693
Qy 326 nLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr----- 342
Db ----- 342
```

```
Db 694 -----ATTGAAGACCTAGAAAACTATGAGCTGCGTGAACCGAT 731
Qy 343 -----ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyr 358
Db 732 TCTGTATACAGCACCATTTGCCCTGAACGGCATGCTATCAATGGCGGCACGGGAGACTG 791
Qy 358 rTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPh 378
Db 792 GCGAACGCAATATCATGACGACGCTGTTTCACCGCTATACGATATT-----CCGCA 842
Qy 378 eSerGlyGlyIle 382
Db 843 TCGCGGAGGCGTT 855
RESULT 26
US-09-938-842A-2606
; Sequence 2606, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2606
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2606

Alignment Scores:
Pred. No.: 0.0366 Length: 1389
Score: 102.50 Matches: 89
Percent Similarity: 33.41% Conservative: 57
Best Local Similarity: 20.37% Mismatches: 164
Query Match: 3.23% Indels: 127
DB: 9 Gaps: 19

US-09-857-669-2 (1-609) x US-09-938-842A-2606 (1-1389)
Qy 89 GlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 58 GGTCTTAATTTGAGAGGAGAACTTGTATGAATTTTCGTCGCTTTTG----- 102
Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 103 -----CGGNAATCAGAAAAAGAT----- 120
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 121 -----CCGTTAAGGATTGTAAGTGTGCTGTGCTGGGGAAGTGTGTTTGA--- 171
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGln-----Gln 164
Db 172 GCATCTTTCAGAAAGCTATGAGGTTTCAGGGTAAGCTTTTCAGATCAGGATATGGAGA 231
Qy 165 ProValGlySerAspPheAspGlnAspSerTrpGlu-----AsnSer 178
Db 232 AGAGCTGGGAGAGCTGTTGATAGAGAAACTGCAGAACATTTCTTCAAGTGATCAATCA 291
Qy 179 LysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAla 194
Db ----- 194
```



```
Db 292 AGGAAGATATCTTGAGGAGATTGATAAGACGCTGTCTTATCTGAAATATGTCGAGGCA 351
QY 195 LysLeuGlyAsnThrArgAlaAlaValAsnPro----- 205
Db 352 AGCTTGGTGATAGGACACTCTATGCTGATGAGATATTGAAACACGGGTTTGTCTTAAC 411
QY 206 -----AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223
Db 412 ATGGTTGATACCGCGCTTCTCTCTTAAGGTTGTGACAAATCTGCAAGAAGCTGTGTGG 471
QY 224 PheGlyAspPheGluIleThrGly-----ThrGlnArgTyrProGluGln 238
Db 472 GATGCTGATATTGTTGTAATGATGCTTCAACTCAACACGTTGAAGAG 531
QY 239 IleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 258
Db 532 ATT---AGTAAGTATTGGAAGAGAGAGATAACCGTTCCGATTATTATCTCTCTCAAG 588
QY 259 AspPheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAsp 278
Db 589 GGTATTGAAACTGCTCTGAA----- 609
QY 279 PheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArg 298
Db 610 -----CCAGTTCCACATATATCACTCAACAAAGATG 642
QY 299 HisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly---LysIle 317
Db 643 ATTCAATCAACCACTGGTGGCGATTGACAATGCTCTGATCTTGGTGACCAACATTT 702
QY 318 AlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMet 337
Db 703 GCTGCTGAATTTTAC-----AACAGGAATATGCCAATGCTAGAACTGTGGAGCT 753
QY 338 AspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsn 357
Db 754 GCTAAATGGAGGAGCCACTAGTAGTCTTAAAGACAACCTCATTTCAAT----- 804
QY 358 TyrTriPThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAla 377
Db 805 GTTTGGGACAATAGTGTATCTTGACACATGAAGTAATGGGAGGTCTC---AAGAATGTC 861
QY 378 PheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAla 397
Db 862 TAGCCCAATTGCA-----GCTGGTATGCTAGCGCGCTCACTAAC 900
QY 398 Glu----- 398
Db 901 GAGAGCGCTACAAGCAAGTCGGTGTATTTTGCTCATTTGTATCTGAGATATTATA 960
QY 399 -----PheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsn 415
Db 961 ACTCATTTACTAGCAGAGAGCCTGACAAACTTGCAGGGCT-----TTGCTAGCTGAC 1014
QY 416 SerHisAlaThrMetLeuThrAla-----SerTyrLysArgGlnLeuLeuAsnAsnVal 433
Db 1015 ACTTATGTGACCTATTATAAAGAGCGTAATGCATGTGACGTCAAATGCTG----- 1065
QY 434 LeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPhe 453
Db 1066 -----GCAAGGGTGAATAAATAAGACAGATGGTGTATGAC 1101
QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1102 ATAAGCGGCAAGGAATGATTACGGGTGTTCTGCACTGGGAGCATTTTAC 1152
```

RESULT 27

```
US-09-956-004-14
: Sequence 14, Application US/09956004
: Patent No. US20072959A1
: GENERAL INFORMATION:
: APPLICANT: Patrick J. Dillon et al.
: TITLE OF INVENTION: Nucleotide Sequences of Escherichia coli Pathogenicity Islands
```

```
: FILE REFERENCE: PB324D1
: CURRENT APPLICATION NUMBER: US/09/956,004
: PRIOR FILING DATE: 2001-09-20
: PRIOR APPLICATION NUMBER: 08/976,259
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/061,953
: PRIOR FILING DATE: 1997-10-14
: PRIOR APPLICATION NUMBER: 60/031,626
: PRIOR FILING DATE: 1996-11-22
: NUMBER OF SEQ ID NOS: 142
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 22671
: TYPE: DNA
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (19750)..(19750)
: OTHER INFORMATION: n equals a, t, g, or c
: NAME/KEY: misc_feature
: LOCATION: (20174)..(20174)
: OTHER INFORMATION: n equals a, t, g, or c
: US-09-956-004-14

Alignment Scores:
Pred. No.: 2.7 Length: 22671
Score: 102.50 Matches: 129
Percent Similarity: 33.76% Conservative: 80
Best Local Similarity: 20.84% Mismatches: 212
Query Match: 3.23% Indels: 199
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-956-004-14 (1-22671)

QY 9 LeuLeuProAlaLeu-----PhePhePheProHisAlaTyrAlaProAlaAla 24
Db 11323 CTCCTCCCTCGCTTTCATTTTCATTTTCCTCGCAGCCATGCTCTCCGGGT 11382
QY 25 Asp-----LeuSerGluAsnLysAlaAlaGly 33
Db 11383 GACCGCAGTCGCAATTCAGCAGCAACAGCAGCAGTTGTTGATGAAACACGCGCCACCGT 11442
QY 34 PheAlaLeuPheLysAsn-----LysSerProAspThrGluSerVal 47
Db 11443 GATGCGCTGGAGCGCAGTCGCGCGCTGACCATCAGCGCGCTCCGGAAACGTCGCGGT 11502
QY 48 LysLeuLysProLysPheProVal-----ArgIleAsp 58
Db 11503 ACTGAAGTCCCTGCTTACGGGTGTCACGCATTTGTCAGTGGGGCCACCGACTGACG 11562
QY 59 ThrGlnAspSerGlu-----IleLysAspMetValGluGluHisLeuProLeuIleThr 76
Db 11563 TCTGCAGAAACCCAGACAGCTGGTGGCACCGCTGGTGAATCAGTGTCTGAATATG----- 11616
QY 77 GlnGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaPro 96
Db 11617 -----ACGGGACTGACCGCG-----GTCAAG 11637
QY 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr 116
Db 11638 GATGCCCTGACGCGACGCTATATACGCGGGGATATATCACCAGCGCGGCTTTCTGACA 11697
QY 117 GluLysAsp---GlyAlaTyrThrValHisIleThr----- 127
Db 11698 GAGCAGGACCTTTCAGGGGGCGCTACTGCACATAACGCTCATGGAAGCAGCGTGCACAA 11757
QY 128 -----ProGlyPro 130
Db 11758 ATCCGGCGGAGCGCTGACCTTCCTCCCGCACCTGAAGATGTTTCCCGGGGATG 11817
QY 131 ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn 150
Db 11818 GAGGGGAAGTCTCTGAACCTGCGGG-----ATATTGAGCAG-GGG--- 11855
```

QY 151 LeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPhe 170
: : : : :
Db 11856 ATGGAGCAGATTAATCTGCTCGGTACGGAGCGGTACAGATTGAATA---TCGCCCGGT 11912
QY 171 AspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGly 190
: : : : :
Db 11913 GACCGTGAGGATGG-----TCGGTGTGACACTGACGGCATTCGCCGAA 11957
QY 191 TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaAsnProAspThrAlaThrVal 210
: : : : :
Db 11958 TGGCCTGTACAGGAGCGGTGGCATCGACAACAGCGGGGAGAGAATACCGTACCGGG 12017
QY 211 AspleuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
: : : : :
Db 12018 CAGTTAAATGGTCTCTTCTTAATCCCTCGGGCTGGCTGACAACTGGTTGTC 12077
QY 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg----PheGlnProGly 249
: : : : :
Db 12078 AGCGGGGAGCGGAGCAGTGTCTTCGGTGTCTACATGATCGAGGAATTTTTCGCCCGGT 12137
QY 250 Thr-----ProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
: : : : :
Db 12138 GTCAGTCTGCCGTATGGCTAT---ACCCTGGTGGATTACACG----- 12176
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp-----Arg 281
: : : : :
Db 12177 -----TATTATCGAGTACTACCTCAGCACCATTATACCCGGGCTGGCGGTGGCGT 12230
QY 282 LeuGlnGlyAspArgValProValLysValSerValThrGluVal-----LysArgHis 299
: : : : :
Db 12231 TCCACGGGAGACCTGACAGCTCAGCGCTGGGACTGTCGATGTCCGTCCGTAACGGG 12290
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
: : : : :
Db 12291 GACATGAAGACAGCAGT-----ACCGAGGTCTGCAGCACCGCATTTATTCAC 12338
QY 320 AspTyrTrpAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
: : : : :
Db 12339 AATTATCTGGAT-----GATGTTCTGCTTCAGGCAGCAGCGGT 12377
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
: : : : :
Db 12378 AAATCTACTTCATTTTCTGTCGGCTGAATCACACACACAAAGTTCTCGGTGTGTCGGA 12437
QY 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
: : : : :
Db 12438 AACTGAATCCGGTATTCACAGG----- 12461
QY 380 GlyGlyIleTyrTrpValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
: : : : :
Db 12462 GGGATGCCCTGGTTC-----GGCGCAGAA-- 12485
QY 400 LeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
: : : : :
Db 12486 AGCCACACCGGAAAGG-----GGAGACCTGCCCTTAATCAGTTCCCGAAATGG 12536
QY 420 MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGly 439
: : : : :
Db 12537 TCGGTGAGTCCAGTTTTCAGCCCGCTCAGGACAGGGTGTGGTGG----- 12584
QY 440 HisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459
: : : : :
Db 12584 ----- 12584
QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
: : : : :
Db 12585 ---CTGACACGCGTTATGCCCAGTGG-----TCACCGGACCGT----- 12620
QY 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSer 499
: : : : :
Db 12621 -----CTTCATGCTGTGGAACTAGGCTC----- 12647

QY 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
: : : : :
Db 12648 -----GGGGGTGAGAGTTTCAGTGCCTGGCTTTTAAGGAGCAGCATATATC 12689
QY 520 GlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeu 539
: : : : :
Db 12690 -----TCCGGTAATAACGGCGGTATCTG-----CGAAATGAGCTGCTGCTGCTCTG 12737
QY 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGly 558
: : : : :
Db 12738 ---TTCCTCCCTGCCATATGTGGGACAGTCCGTGCAGTGCACATGGAGCGGCG 12791
RESULT 28
US-09-917-800A-1505
: Sequence 1505, Application US/059117800A
: Patent No. US20020119462A1
: GENERAL INFORMATION:
: APPLICANT: Mendrick, Donna
: APPLICANT: Porter, Mark
: APPLICANT: Johnson, Kory
: APPLICANT: Castile, Arthur
: APPLICANT: Elashoff, Michael
: APPLICANT: Gene Logic, Inc.
: TITLE OF INVENTION: Molecular Toxicology Modeling
: FILE REFERENCE: 44921-5038-US
: CURRENT APPLICATION NUMBER: US/09/917,800A
: CURRENT FILING DATE: 2001-07-31
: PRIOR APPLICATION NUMBER: US 60/222,040
: PRIOR FILING DATE: 2000-07-31
: PRIOR APPLICATION NUMBER: US 60/222,880
: PRIOR FILING DATE: 2000-11-02
: PRIOR APPLICATION NUMBER: US 60/290,029
: PRIOR FILING DATE: 2001-05-11
: PRIOR APPLICATION NUMBER: US 60/290,645
: PRIOR FILING DATE: 2001-05-15
: PRIOR APPLICATION NUMBER: US 60/292,336
: PRIOR FILING DATE: 2001-05-22
: PRIOR APPLICATION NUMBER: US 60/295,798
: PRIOR FILING DATE: 2001-06-06
: PRIOR APPLICATION NUMBER: US 60/297,457
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/298,884
: PRIOR FILING DATE: 2001-06-19
: PRIOR APPLICATION NUMBER: US 60/303,459
: PRIOR FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 1740
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1505
: LENGTH: 15231
: TYPE: DNA
: ORGANISM: Rattus norvegicus
: FEATURE:
: OTHER INFORMATION: Genbank Accession No. US20020119462A1 X59601
US-09-917-800A-1505

Alignment Scores:
Pred. No.: 1-9 Length: 15231
Score: 101.50 Matches: 113
Percent Similarity: 34.53% Conservative: 80
Best Local Similarity: 20.21% Mismatches: 226
Query Match: 3.20% Indels: 140
DB: 10 Gaps: 28

US-09-857-669-2 (1-609) x US-09-917-800A-1505 (1-15231)

QY 64 IleLysAspMetValGluGluHis---LeuProLeuIleThrGlnGlnGluVal 82
: : : : :
Db 12574 ATCATTCACCTCGAGAGAGCCACCGCTGCCTGTGGAAGTGGCGCTATAAGCGTGGTCTC 12633
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMet 102
: : : : :
Db 12634 TTTGATGAGGAGATGAACGAGATCCTGCTGACCCCTCAGATGAC----- 12678

```
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
Db 12679 -----ACCAAGGCTTCTTCGACCCAAAC-----ACCGAGGAAACCTCACATAC 12723
QY 123 -----ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly--- 138
Db 12724 CTGACGTGATGGAGCCCTGTATCACTAGACCCCGACACTGGCCTGTCTCTGCCCGCTG 12783
QY 138 ----- 138
Db 12784 AAGGAAAGAGCGGAGCGGAAGACGCTCTCCAAAGTCTCAGTGCACGCGCGCGTG 12843
QY 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAla 158
Db 12844 GTGATTGTGGACCCCTGAGCGGCGGAAGAGATGTCACTGTATGAGGCTACCGCAAGGCG 12903
QY 159 LeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSer 178
Db 12904 CTATAGACCAACAGACATACCTGGAGTTGTGAGAGCAGGATGCGAGTGGGAAGAAATC 12963
QY 179 LysThrSerValLeuGlyAlaValThrArgLys----- 189
Db 12964 ACCATCTCTTCTCGGAGCGGCTGCTCAATCTATGATCATCGACCGCGCTCTGGCGCG 13023
QY 190 -----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 13024 CAGTATGACATTTGTCAGCGGCATCACCAGAACCTCATTTGACCGCTCAGCACCTGGACCAG 13083
QY 206 AspThrAla---ThrValAspLeuAsnValValAsp-----SerGlyArgProIle 222
Db 13084 TACCGCGCTGCACACATTTCTATCACCAGGTTTGCACGACATGCTCTCAGGCAACGCTGGT 13143
QY 223 AlaPhe-----GlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 13144 GGCTTCGCTCCCGCTCTCTCTGCTGGGCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13197
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 13198 TCTGCTGCTCCAGGACCCAG-----CTAGCCTCTCTGG 13230
QY 261 GlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 13231 TCTGATCCTACTAGGAGACTGGCCGAGTGGCCGAGATC-----CTAGACACAGAG 13281
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 13282 ACTCTGGAG-----AAGGTGTCATCATCAGAGCGCTGCACCGGCGGCATC 13323
QY 301 LeuGlu-----ThrGlyIleArg---LeuAspSerGluTyrGlyLeuGlyGlyLys 316
Db 13324 CTGGTAGACAACATCACTGGCCAGCGTTGCTGGAGGCACAGCGCTGCACCGGGGCGATC 13383
QY 317 Ile-----AlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIle 330
Db 13384 ATTGACCCCGACACTGGTGGAGCGTCTCCCGTCACTGAGCGTGTCAACAAGGCGCTGGTG 13443
QY 331 GlySerValValTrpAspMetAspLysTyrGlnThrThrLeuAlaAlaGlyIleSerGln 350
Db 13444 GACAAGATCATGGTAGACCGTATCAATCTGCCCGCAGAGCGCTCTCTGGGTTTGAGGAC 13503
QY 351 ProArgAsn-----TyrArgGlyAsnTyrTrpThr 360
Db 13504 CCAGGCACCAAGACCAAGATGTCAGTGCACCGCGCTGGAAGAGGCG-----TGGCTT 13557
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 13558 -----TACTACAGGCGGCGCGGTTTCTCGAGGTGCAGTACCTACCGGGT 13605
QY 381 GlyIleTrp-----TyrValArgAspArgAla 389
Db 13606 GGCTGTATGAGCTGACACACCTGGCGGTGTGTCTTGTATGAGCGCTTGCACACGCGGC 13665
QY 390 GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAlaGluArg 404
```

```
Db 13666 ACTGTGATGCCGCGACAGCCAGAGCTGCGTATGTCAGTGCCTACTCAAGTACCTC 13725
QY 405 LysIleProGlySerAspIleAspLeuGlyAsnSerHisAla-----ThrMetLeu 421
Db 13726 ACGTGCCCAAGACCAAGCTCAAGATCTCTTACAGGACGCTGTGGATCGGAGCATGGT 13785
QY 422 ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyr 441
Db 13786 GAGGAGGCGACAGGGCTGAGGCTCTCGAAGCCGCGCAGTCCACAGGGCTACTAC 13845
QY 442 LeuAspGlyLysIle-----GlyThrThrLeuGlyThrPheLeuSerSerThrAla 458
Db 13846 AGCCCGTACAGTGTCTGAGTGTCTGCTCTACTGCTGTCTTCCACGACCTGGTTCA----- 13899
QY 459 LeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGly 478
Db 13900 -----CGCACCGGCTCCAGGGCGGCG-----TCCCGTCGTGGC 13932
QY 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValPro 498
Db 13933 AGCTTTGATGCCACTGGC---TCTGGCTTCTCCATGACCTTTTCTTCTCTCTCTCTCT 13989
QY 499 Ser-----GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGlu 515
Db 13990 TCCTCAGGCTATGGCGCGCTATGCTCAGGCGCTTCAAGGCTCTCTTGGGGCGCTGAG 14049
QY 516 LeuAspSerIleGlyLeuAla-GlyProAsnGlySerValLeuProGluArgAla 533
Db 14050 -----TCTGCAGTGGGCTGATCCCGCAGCGCTGATCTCTCTCTCTCTCTCTCT 14095

RESULT 29
US-09-841-132-170
; Sequence 170, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 170
; LENGTH: 2949
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-170

Alignment Scores:
Pred. No.: 0.173 Length: 2949
Score: 101.00 Matches: 133
Percent Similarity: 32.85% Conservative: 70
Best Local Similarity: 21.52% Mismatches: 237
Query Match: 3.18% Indels: 178
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-841-132-170 (1-2949)
QY 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
Db 1003 TTTCTTTAGTAGCAATGTAGCTGCTGGGAAAGGGGAGCTATTATGCAAAAAGCTCTCG 1062
QY 128 Pro-----GlyPro-----ArgThrLysIleAlaAsnValGlyValAlaIle 141
Db 1063 GTTCCTNACTGTGGCCCTGTACAAATTTTAAAGGAATATCGCTAATGATGGTGGAGCGATT 1122
QY 142 -----LeuGlyAspIleLeuSerAsp 148
Db 1123 TATTAGGAGAATCTGGAGAGCTCAGTTTATCTGCTGATTTATGGAGATATATTATTTGAT 1182
```

```
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
    |||||
Db 1183 GGAATCTTAA--AGACAGCAAGAGAAATGCTGCCGATGTTAATGGCGTA 1233

Qy 169 AspPheAspGlnAspSer-----TrpGluAsnSerLysThrSerValLeuGlyAla 185
    :|||
Db 1234 ACTGTCTCCTCAACAGCCATTTTCGATGGGATCGGAGGGAATAACACATTAAGAGCT 1293

Qy 186 ValThrArgLysGlyTyrProLeu-----AlaLysLeuGlyAsn 198
    :|||
Db 1294 -----AAGAGGCGCATCAGATCTCTTAAATGATCCCATCGAGATGGCAACGGAAT 1347

Qy 199 ThrArgAlaAla-----ValAsnPro----- 205
    :|||
Db 1348 AACAGCCAGCGCAGCTTCCAAACTTCTAATAATTAACGATGGTGAAGGATACACAGG 1407

Qy 206 -----AspThrAlaThrValAspLeuAsnValValValAspSerGly 219
    :|||
Db 1408 GATATTGTTTTCGTAATGGAAGCAGTACTTTGTACCAAAATGTTACGATAGCAAGGA 1467

Qy 220 Arg-----ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
    :|||
Db 1468 AGGATTGTTCTCGTGAAGGCAAAATATCAGTGAATCTCTAAGTCACACAGGTGG 1527

Qy 233 GlnArgTyrProGlu-----GlnIleValSerGlyLeuAlaArgPheGln 247
    :|||
Db 1528 AGTCTGTATGGAAGCTGGGAGTACATTTGGATTGTAAGTCCACCAACCACCAACAG 1587

Qy 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu----- 265
    :|||
Db 1588 CCTCTCGCGTAAATCAGTTGATCAGCTTTTCCAAATTCGATTTCTCTTTCTTTG 1647

Qy 266 -----GlnAsnGlyHis--- 269
    :|||
Db 1648 TTAGCAACAATGACGTACGATCCTCTTACCAATCTCTCCAGCGCAAGATCTCATCT 1707

Qy 270 -----TyrSerGlyAlaSerValGlnAla 277
    :|||
Db 1708 GCAGTCATTGGTAGCACAACTGCTGTTCTGTACAAATAGTGGGCGCTATCTTTTTCAG 1767

Qy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
    :|||
Db 1768 GATTGGATGATACAGCTTATAGGTATAGGTATGATGGCTAGGTTCTTAATCAAAAAATCAAT 1827

Qy 298 ArgHisLysLeuGluThrGly-----IleArgLeu 307
    :|||
Db 1828 GTCCTGAAATTACAGTTAGGCACTAAGCCCGAGCTAATGCCCCATCAGATTTGACTCTA 1887

Qy 308 AspSerGlu-----TyrGlyLeuGlyGly-----LysIleAlaTyrAspTyrTyr 322
    :|||
Db 1888 GGAATGAGATGCCCTAAGTATGGCTATCAAGGAGCTGGAAGCTTGGCTGGGATCCTAAT 1947

Qy 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThr 342
    :|||
Db 1948 ACAGCAATAATGTCCTTACTACTGAAAGCTACATGAGCTAAGAACTGGGTATAATCCT 2007

Qy 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
    :|||
Db 2008 GGGCTGAGCGAGTAGCTTCTTGGTCCAAAT-----AGTTTATGGGATCCCAT 2058

Qy 363 Val-----SerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
    :|||
Db 2059 TTAGATATACGATCTCGCATTCAGCAATTCAGCAAGCTGGATGGCGCTCTTATTGT 2118

Qy 380 GlyGlyIleTyr-----TyrValArgAspArgAlaGlyIleAsp 392
    :|||
Db 2119 CGAGATTATCGGTTCTGGAGTTTCGAATTTCTTCTATCATGACCGCGATGCTTATAGT 2178

Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
    :|||
Db 2179 CAG-----GGATATCGGTATATTAGTGGGTTATTCTCTTA---GGAGCAAACTCCTAC 2229
```

```
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
    :|||
Db 2230 TTTGGATCATCGATGTTTGGTCTAGCATTTACCGAAGTATTTGGTAGATCTAAAGATTAT 2289

Qy 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle-----Gly 447
    :|||
Db 2290 GTAGTGTGTGCTTCCCAATCATCGTTCGATAGATCGGTTTATCTATCTACCCAACAA 2349

Qy 448 ThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGly 467
    :|||
Db 2350 GCTTTATGTCGATCTATTTGTCGAGATGCGTTTATC-----CGTGCTAGC 2397

Qy 468 TyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGly 487
    :|||
Db 2398 TACGGGTTTGGGAATCAGCATATGAAA-----ACCTCA 2430

Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
    :|||
Db 2431 TATACATTTCGACAGGAG----- 2448

Qy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
    :|||
Db 2449 ---ACGATGTTCTG-----TGGGATAATAACTCTCTGCTGGAGAGATTGGAGCG 2496

Qy 528 ValLeuPro-----GluArgAlaLeuValGlySerLeuGluTyrGlnLeu 543
    :|||
Db 2497 GGATTACCGATTTGTCATTACTCCATCTAAGCTCTATTTGAATGAGTTCGCT----- 2547

Qy 544 PropheThrArgThrLeuSerGlyAlaValPheHisAspMet-----GlyAsp 559
    :|||
Db 2548 CCTTTGTCGCAAGCTGAGTTTCTTATCCGATCATGAATCTTTTACAGAGGAGCGCAT 2607

Qy 560 AlaAlaAlaAsnPheLysArgMetLysLysLeuLysHisGlySer---GlyLeuGlyValArg 578
    :|||
Db 2608 CAAGCTCGGGCATTCAGAGCGGACATCTCTAAATATCATCATGTCCTGTTGGAGTGAAG 2667

Qy 579 Trp-----PheSerProLeuAlaPro-----PheSerPheAspIleAlaTyr 592
    :|||
Db 2668 TTTGATCGATGTTCTAGTACACATCCTAATAAATATAGCTTTATGGCGCTTAT 2721

RESULT 30
US-09-841-132-182
; Sequence 182, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 182
; LENGTH: 3021
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-182

Alignment Scores:
Pred. No.: 0.179 Length: 3021
Score: 101.00 Matches: 133
Percent Similarity: 32.85% Conservative: 70
Best local Similarity: 21.52% Mismatches: 237
Query Match: 3.18% Indels: 178
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-841-132-182 (1-3021)

Qy 108 TyrPheSerSorLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
    :|||
    :|||
```

Db 1075 TTCTTTAGTACCAATGTAGTGTGGGAAAGGGGGAGCTATTTATGCCAAAAAGCTCTCG 1134
Qy 128 Pro-----GlyPro-----ArgThrLysIleAlaAsnValGlyValAlaIle 141
Db 1135 GTTCCTAACTGTGGCCCTGTACAAATTTTAAAGGAATATCGCTATATGTTGGAGCGCAT 1194
Qy 142 -----LeuGlyAspIleLeuSerAsp 148
Db 1195 TATTTAGGAGAACTGGAGAGCTCAGTTTATCTGCTGATTGATGGAGATATATTTTCGAT 1254
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 1255 GGGAACTCTTAA-----AGAACAGCAAGAGAGATGCTGCCGATGTAAATGGCGTA 1305
Qy 169 AspPheAspGlnAspSer-----TrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1306 ACTGTCTCTCACAAAGCCATTTTCATGGGATCGGGAGGGAATAAGCAGATTAAAGAGCT 1365
Qy 186 ValThrArgLysGlyTyrProLeu-----AlaLysLeuGlyAsn 198
Db 1366 -----AAAGCAGGGCATCAGATTCTCTTAAATGATCCCATCGAGATGGCAACCGAAAT 1419
Qy 199 ThrArgAlaAla-----ValAsnPro----- 205
Db 1420 AACCAGCCAGCGCAGCTCTCCAACTTCTAAAAATTAACGATGGTGAAGGATACACAGGG 1479
Qy 206 -----AspThrAlaThrValAspLeuAsnValValAspSerGly 219
Db 1480 GATATTGTTTGTCTAATGAGCAGTACTTTGTACCAAAATGTACGATAGACCAAGGA 1539
Qy 220 Arg-----ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 1540 AGGATTGTTCTCGTGAAGGCAAAATATATCACTGAATCTCTAAGTCAGACAGGTGG 1599
Qy 233 GlnArgTyrProGlu-----GlnIleValSerGlyLeuAlaArgPheGln 247
Db 1600 AGCTGTATATGGAAGTGGGAGTACATGGATTTTCTAATCCACCAACCAACACAG 1659
Qy 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu----- 265
Db 1660 CCTCTCGCGCTAATCAGTTGATCAGCTTTCCTCAATCTGCAATTCCTCTCTCTCTCTCT 1719
Qy 266 -----GlnAsnGlyHis----- 269
Db 1720 TTAGCAAAACAAATGAGTACGAATCTCTACCAATCTCCAGCGCAAGATTCATCCT 1779
Qy 270 -----TyrSerGlyAlaSerValGlnAla 277
Db 1780 GCAGTCATTTGGTACCAACTGCTGGTCTGTACAAATTAGTGGGCGCTATCTTTTGGAG 1839
Qy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
Db 1840 GATTGTGATGATCAGCTTATGATAGTATGATGATGATGATGATGATGATGATGATGAT 1899
Qy 298 ArgHisLysLeuGluThrGly-----IleArgLeu 307
Db 1900 GTCTCGAAATACAGTTAGGAGTAAAGCCAGTAAAGTCCCATCAGATTTGACTCTA 1959
Qy 308 AspSerGlu-----TyrGlyLeuGlyGly-----LysIleAlaTyrAspTyrTyr 322
Db 1960 GGAATGAGATGCCATAGATAGTGGTATCAAGGAAGCTGGAAGCTTGGTGGGATCCTAAT 2019
Qy 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr 342
Db 2020 ACAGCAATAATGTCCTTATCTCTGNAAGCTACATGGACTAAAGTGGGTATAATCCT 2079
Qy 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
Db 2080 GGGCTGAGCAGTAGCTTCTTTGTTCCAAAT-----AGTTTATGGGATCCATT 2130
Qy 363 Val-----SerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
Db 2131 TTAGATACGATCTGGCGCATTCAGCAATTCAGCAAGTGTGGATGGGCGCTCTATTGT 2190

Qy 380 GlyGlyIleTrp-----TyrValArgAspArgAlaGlyIleAsp 392
Db 2191 CGAGGATATGGGTTTCTGGAGTTTCGAATTTCTTCTATCATGACCGCGATTTAGGT 2250
Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
Db 2251 CAG-----GGATATCGTATATTTAGTGGGGTTATTCCCTTA---GGAGCAAACTCCTAC 2301
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
Db 2302 TTTGGATCATCGATGTTGGTCTAGCATTTACCGAAGTATTTGGTAGATCTAAAGATTAT 2361
Qy 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle-----Gly 447
Db 2362 GTAGTGTGCTGTTCCATCATCATGCTTGCATAGCATCCGTTTATCTATCTACCCAACAA 2421
Qy 448 ThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGly 467
Db 2422 GCTTTATGTGATCCTATTTGTTTCGGAGATCGGTTTATC-----CGTGCTAGC 2469
Qy 468 TyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGly 487
Db 2470 TACGGGTTTGGGAATCAGCATATGAAA-----ACCTCA 2502
Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
Db 2503 TATACATTTCCAGAGGAG----- 2520
Qy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
Db 2521 ---AGCGATGTTCTG-----TGGGATAAATAACTGTCTGCTGGAGAGATTGGAGCG 2568
Qy 528 ValLeuPro-----GluArgAlaLeuValGlySerLeuGluTyrGlnLeu 543
Db 2569 GGATTACCGATGTGATTTACTCTCAAGCTCTATTGGAATGAGTTGCGT----- 2619
Qy 544 ProPheThrArgThrLeuSerGlyAlaValPheHisAspMet-----GlyAsp 559
Db 2620 CCTTTCGTGCAAGCTGAGTTTCTTATCCCGATCATGAATCTTTTACAGAGGAAGCGAT 2679
Qy 560 AlaAlaAlaAsnPheLysArgMetLysLysHisGlySer-----GlyLeuGlyValArg 578
Db 2680 CAAGCTGGGCGATTCAAGAGCGGACATCTCTCAAAATCTATCAGTTCTCTGTGGAGTGAAG 2739
Qy 579 Trp-----PheSerProLeuAlaPro-----PheSerPheAspIleAlaTyr 592
Db 2740 TTTGATCGATGTTCTTAGTACACATCTCTAAATAATATAGCTTTATGGCGGCTTAT 2793

RESULT 31

US-09-070-927A-527

Sequence 527, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

Qy 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
||||| : : : : :
Db 3728 GTTCCATTAAAGATCTCGATGATCTCGAGCGGCACGACGAGTAAAGCAGCAGAA-- 3784
Qy 517 AspSerIleGlyLeuAlaGly-----ProAsnGlySerVal 528
||| ||||| : : : : :
Db 3785 ---GACGCGAGCAAGCTGGAGCAGATAGAAAGCCGGAAGTGGAAACCGAGCGGTTTAGTG 3841
Qy 529 LeuProGluArgAlaLeuLeuValGlySerLeu 539
||||| : : : : :
Db 3842 ACTCCAGAGAAAGCGGCGAGTGGATGGCTTG 3874

RESULT 32

US-09-815-242-6700
: Sequence 6700, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6700
: LENGTH: 2820
: TYPE: DNA
: ORGANISM: Enterococcus faecalis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(2820)
US-09-815-242-6700

Alignment Scores:
Pred. No.: 0.184 Length: 2820
Score: 100.50 Matches: 140
Percent Similarity: 29.97% Conservative: 89
Best Local Similarity: 18.32% Mismatches: 248
Query Match: 3.17% Indels: 287
DB: 10 Gaps: 36

US-09-857-669-2 (1-609) x US-09-815-242-6700 (1-2820)

Qy 42 ProAspThrGluSerValLys---LeuLysProLysPheProValArgIleAspThrGln 60
||||| : : : : :
Db 217 CCAGATGTTGACAGCATGATGGTTTAAAGTCCAGCTATTTCATTTGATCAAAAGCAGACA 276
Qy 61 Asp-----SerGluIleLysAspMetVal----- 68
: : : : :
Db 277 AGTAAACCGCGCTCAACGGTGGGAGCGGTACGGAAATCAATGATTATTACGATTG 336

Qy 69 -----GluGluHisLeuProLeuIleThrGln 77
: : : : :
Db 337 TTATTTCTCGGTAGCTCATCAATTTGCTCAACGATCATATTGAAATACCAGTCAG 396
Qy 78 GlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAsp 97
: : : : :
Db 397 TCTGTTCAACAGATGGTTGATAA-----GTCTTAGAGCTACCAGAG 438
Qy 98 AsnValLys-----ThrMetLeuArgSerLysGlyTyrPheSerSer 111
: : : : :
Db 439 CGAACAAAAATTCAAATTTTGGCGCCTGTTGCTGCTCAAGAAAAAGCCCAATATAAAAA 498
Qy 112 LysValSerLeuThrGluLysAspGly-----AlaTyr 122
: : : : :
Db 499 GTGTTTCAATGATCCAAACGTGAAGGTATGTCAGAAATCGGGTGGATGGTGAACCTAT 558
Qy 123 ThrValHisIleThrPro----- 128
: : : : :
Db 559 GATGTTAGTAGGCAACCAAGAACTTGAATAAATAAATAAATATATATGCGATTGTGATT 618
Qy 129 -----GlyProArgThrLysIleAlaAsn----- 136
: : : : :
Db 619 GACGCTATCGTTGAAAGAGGTATTGCTCTCGCTGTTGATTGCTTTGAACGGCT 678
Qy 137 -----ValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu--- 151
: : : : :
Db 679 TTGCGTTTACCTGAAGGTACGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
Qy 152 AlaGluTyrTyr-----ArgAsnAlaLeu 159
: : : : :
Db 739 AGTCAACATTATGCTTCTGCTTATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798
Qy 160 GluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLys 179
: : : : :
Db 799 TTTTCATTAAATGACCACTTTGGCGCTTGCTCCGATGCTGATGCTGATGCTGATGCTG 843
Qy 180 ThrSerValLeuAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199
: : : : :
Db 844 -----TTAGGT-----GTTAATTTAGAGTGGAC 867
Qy 200 ArgAlaAlaValAsnProAsp---ThrAlaThrValAspLeuAsnValValAspSer 218
: : : : :
Db 868 AAGACTTAGTGTATCCAGATCCCACTAAACCTTGAGAGAAGGAGGACCATGTT---CCG 924
Qy 219 GlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGln 238
: : : : :
Db 925 TGAACCCCTATTACT-----TCCCAATATTATCCACAAATG 960
Qy 239 IleValSerGlyLeuAlaArgPhe-----GlnProGlyThrProTyrAspLeuAspLeu 256
: : : : :
Db 961 TTGGAGCAACGACGCTACCGATTTTGGGATGATGATGATGATGATGATGATGATGATGAT 1017
Qy 257 LeuLeuAspPheGlnGlnAlaLeu-----GluGlnAsnGlyHisTyrSer 271
: : : : :
Db 1018 CTGCAGACCAACAAGAAATATTATTTAAATGGTCTGCGGAGAGAAATCTTCAATTT--- 1074
Qy 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysVal 291
: : : : :
Db 1075 -----CATTATGAAATGACTTTGGTGGTGTCTGCTGATGTGGAAGTACCATTGGAAGGT 1128
Qy 292 SerValThrGluLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyr 311
: : : : :
Db 1129 ATTTTAAAAATATTAAACGAGGTATCATGAAACAAATACGAGATTTTACACGAGAT--- 1185
Qy 312 GlyLeuGlyLysIleAlaTyrAspTyrTrpAsnLeuPheAsnLysGlyTyrIleGly 331
: : : : :
Db 1185 ----- 1185
Qy 332 SerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro 351
: : : : :
Db 1186 -----CAATGCGGTTATACATGACAGAAATGACTTGTGCGAAGCTGCA--- 1230

Qy	352	ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln	371
Db	1231	-----GGGTTCGTTGAATCGCGAAGCTTTAGCTGTAAAAATCAATGGCACGCACATTGGT	1287
Qy	372	AsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyr-----	384
Db	1288	GAAGTCAGTGAATAGCAATAAAAAATCGGTCGCAATTTTTTGAAGGTGTCTTTATCT	1347
Qy	385	-----ValArgAspArgAlaGlyIle	391
Db	1348	GAACAAGAAATAAGATTGCTCGGCCAAATTTTAAAGAAGTCGAGATCGGTTAACCTTC	1407
Qy	392	AspAlaArgIeuGlyAlaGluPheLeuAlaGluArgLysIleProGlySerAspIle	411
Db	1408	TTAAAAAATGTTGGTTAGATTACTTTAACTTTAAGTCGA-----GCTGCTGGTACACTT	1461
Qy	412	AspLeuGlyAsnSerHis-----AlaThrMetLeuThrAlaSerTrpLysArg	427
Db	1462	TCTGGCGGAGACCAACAAGATTCGCTTAGCAACACAAATTTGGATCAAAAC-----	1512
Qy	428	GlnLeuLeuAsnValLeuHis-----	436
Db	1513	-----TTATCAGCGCTCTTTACATTTTAGATGAACCATCAATTTGGTTGCATCAACGA	1566
Qy	437	GluAsnGlyHisTyrLeuAspGly-----LysIleGlyThrThrLeu---	450
Db	1567	GATAACGACCGCTTGATTGACTCATTTGAAAAAATCGGGACTTAGGCAATACATTAATT	1626
Qy	451	-----GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla	464
Db	1627	GTGCTGGAACACGATGAACATAAATGATGCTTCGGATTACTTGATGTATGGGACCT	1686
Qy	465	ArgAlaGlyTyrPhe-----PheThrProGluAsnLysLysLeu	477
Db	1687	GGCGCTCGACATCTGGCGCGTGAAATTTAGCCGCTGTACCGCGAAGAGTTGCTAAA	1746
Qy	478	GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspVal	497
Db	1747	AATCCGCAATTCATTGACTGGACAA-----TATCTTTCTGGGAAAAAGTCATTCTGTGA	1800
Qy	498	ProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAsp	517
Db	1801	CCAAAGAAGCGTCGCAAGGTAATGGCAAGCGATCAAAAGTCACGTGCGCAGTGAAT	1860
Qy	518	SerIle-----GlyLeu	521
Db	1861	AATTTAAAAAATGTCAGTGTGAATTTCCGCTAGCTGAATTTGTGCTGTACAGGCGTT	1920
Qy	522	AlaGlyProAsnGlySerValLeuProGluArgAlaLeu-----	535
Db	1921	TCAGGTTTCAGGGAAGTAGTACTTTAGTCAATCAAAATTTTGAAAAAAGCACTGGCCCAAAA	1980
Qy	535	-----	535
Db	1981	TTAATGCTAATTCCAATAAAGCTGGGAACACAAAAGTATCACTGGTTATGAAGCCATT	2040
Qy	536	-----ValGlySerLeuGluTyrGlnLeuProPhe	545
Db	2041	GAAGAAGATTGTCACATTGATCAAAAGTCCAAATTTGTGCGAAGCCCAAGAAGTAATCAGCG	2100
Qy	546	ThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLys	565
Db	2101	ACTTATACT-----AGTCTTTTGTGATGACATTCGCGATTTGTTGCCCAACAAT	2151
Qy	566	ArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaPro	585
Db	2152	GAAGCCCAAG-----GTTCTGTGTTATAAAAAA-----GGCCGC	2184
Qy	586	PheSerPheAspIleAlaTyrGlyHisSerAspLys-----Lys	598
Db	2185	TTTAGTTTTAACGTCAAAGGTGGCCGTTGTGAAGCTGTGTCGGGGGATGATATCATCAAA	2244
Qy	599	IleArgTrpHis	602

QY	88	ThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGly	107
Db	11708	GCGGGCTTTTAAACCAATATTATTACCAGAT-----GCTGAACGCCAAGAT	11664
QY	108	TyrPheSerLysValSerLeuThrGlnLysAspGlyAlaTyrrThrValHisIleThr	127
Db	11563	GTCCGCACAGACATGAATTGCTTAAGTGAGTCCTCGGGTCTATATTTCGGTAATTAGCT	11604
QY	128	ProGlyProArghThrLysIleAlaAsnValcLyValAlaIleLeuGlyAspIleLeuSer	147
Db	11603	-----TTTGATACG	11595
QY	148	AspGlyAsnLeuAlaGluTyrrTyrArgAsnAlaLeuGluAsnTrpGlnInProValcGly	167
Db	11594	CAGGGACGCTTTTGATAATTTTTACAACGACAGCAGAACACACTTTTTACGACGGCGTCAA	11535
QY	168	SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr	187
Db	11534	TTTAGTTATCGAATTGACGACGACGCANTAAATAC-----	11499
QY	188	ArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaLaValaValAsnProAspThr	207
Db	11498	-----AATGCTCAATGCTCGCATGAT-----	11478
QY	208	AlaThrValAspLeuAsnValVal-----ValAspSerGlyArgProIleAlaPhe	224
Db	11477	-GATTTACGGATCATTCGTTTCATTAATTGAAGCAGGG-----	11442
QY	225	GlyAspPheGluIleThrGlyThrGlnArgTyrProGlnGlnIleValSerGlyLeuAla	244
Db	11441	GGTCATTTTAAA-----ACAGATCAGTATGATCAAGAANAATAAAAAATTAGGCAG	11391
QY	245	ArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAlaLeu	264
Db	11390	AGTTTATGAAACCTCTATGAAAGAT---AATATTCTAATTGATTTT-----	11346
QY	265	GluGlnAsnGlyHisTyrSerGlyValAsaSerValGlnAlaAspPheAspArgLeuGlnGly	284
Db	11345	-----TAGCATAGTAAATCAAAACA-----	11325
QY	285	AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGly	304
Db	11324	-----CAGTCTTCAGAAACTCTTTGTTTATATATGACTTAATTACCGTG	11280
QY	305	IleArgLeuAspSerGluTyrrGlyLeuGlyLysIleAlaTyrrAspTyrrAsnLeu	324
Db	11279	GGTTATCTATACAAGAATTTGGGATTTTCAGCACACTTTCAGCATCATTTACCAATTA	11220
QY	325	PheAsnLysGlyTyrrIleGlySerValValTrpAspMetAspLysTyrrGluThrThrLeu	344
Db	11219	ATAGATGATGGCTATATTTCGTGAT-----GATTTACCTCTTTATCAACCAAGTTT	11169
QY	345	AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSer	364
Db	11168	AAT-----CATCAACAANAATAACGAAACAAC---GGACACACTTAATATTATT	11121
QY	365	TyrAsnArgSerThrThrGlnAsnLeuGluLys-----ArgAlaPheSerGlycylIle	382
Db	11120	GAATCTTATTAAAGCATTGTCCATTTTAGTGAAGTGGGTATGGCGACAGCAGTAGTATT	11061
QY	383	TrpTyrrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu	402
Db	11060	GATTTTGGCGAAACA-----GTTCAAGCAAGCAAGCGCTTTT-----AAC	11019
QY	403	GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr	422
Db	11018	TGCTATGATTAAATGGATCGCCCGTGGATAAATAATCAATCTGCAGCTAGTATTCGGATA	10959
QY	423	AlaSer	424
Db	10958	GCGGCT	10953

```
Db 653 CAAACCGTGGCCCTTGTGACCAAGCAATATCAAAAGGCGAAACCAACCAAAATTTACTGGG--- 709
Qy 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 710 ---GAAACAACTCTGATAAAGACAGGATACTGGAGATAAACAACAGGCGGCA 766
Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly----- 120
Db 767 GTCTTTGAGGGAGCTGAATATACCTTGTCTTACTGCTTAAGGATGGAAAGCGGTAAATGG 826
Qy 121 -----AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly 138
Db 827 ACTGAAGCTTTAAACAGCAATTTGTTAAAGGNAACCAAGCTTCTGATCAACAGTGACT 886
Qy 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla----- 152
Db 887 CTGGCT-----TTAGATGAAGAAGATCAAGCTGCCGTTAAACACTTTAGCC 931
Qy 153 -----GluTyrTyrArgAsnAlaLeuGluAsnTyrPheGln-----ProValGly 167
Db 932 ATCAACGAGTACTAC-----TGGCAAGAAACCAACCAACCACTGAAGGA 973
Qy 168 SerAspPheAspGlnAspSerTyr-----GluAsnSerLys 179
Db 974 TATACCTTCGATGAACAGGATATCTGTTCYATCAAAAARGTGWATGATACGAAAAA 1033
Qy 180 ThrSerValLeu-----GlyAlaValThrArgLysGlyTyrPro 192
Db 1034 AATCGCGTATTACYGAGATGTTACGGCAAGAACAGATTTTCGTTTGGTTTGAT 1093
Qy 193 LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeu 212
Db 1094 TTCCTCAAAATTCGCGGATTCRCYGMTGC-----ACTGCCGAACACTGGGTTT 1141
Qy 213 AsnValValAspSerGlyArgPheGluAlaPheGlyAspPheGluIleThrGlyThr 232
Db 1142 ATGACTTAACCTTTAAAGTGTCCGCAATRGAA---GGGACCAATGAGATCACAGTGT 1198
Qy 233 Gln-----ArgTyrProGluGlnIle---ValSerGlyLeuAlaArgPhe 246
Db 1199 GAAGATRAAGCAATCACAGCTTATAATAGCAATTAGGCTTTGATGGCTACGGAAAGTTT 1258
Qy 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 1259 GAA---AATCTTCCCTTATGGG---GATTATTACTTTGAAGAAGTAGAAGCGCCAGGA 1312
Qy 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArg 286
Db 1313 TTTCAAAAATTTACGCCATTAGCAATCCGTTCTTACTTTTAAAGAAACCAACAGACITTT 1372
Qy 287 ValProValLys-----ValSerValThrGluVal---LysArgHisLysLeuGluThr 303
Db 1373 CTGAAGAGTGAGTATGCTCTTTACCATTTACGCAAGGCGCAACCAACCAATTAACG 1432
Qy 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsn 323
Db 1433 -----GTGACCGCTTCCTTTACGAAAG 1453
Qy 324 LeuPheAsnLysGlyTyrIleGlySer-----ValValTyrAspMetAspLys 339
Db 1454 CTGACGAACAAAGCATTTCTGTAGTTTGAACCGTTTGATGCTCTATGATTTGCTGAG 1513
Qy 340 TyrGlu-----ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 1514 GAAGAAGATAGTTTGACTTCTCTTCCAACTTGGGAAGGCGGAAATATAAATAATGAATACC 1573
Qy 357 AsnTyrTyrThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 1574 CTGTATTTTACCGAGCTAGTTGATAAATTTGAGATATATACCTTGCATGAAATCAAGAAC 1633
Qy 377 AlaPheSerGlyGlyIleTyrTyrValArgAspArgAla---GlyIleAspAlaArgLeu 395
Db 1634 -----TGGTATCTCGTCAAGCCATTGATGTGGAACCCACAAA 1675
```

```
Qy 396 GlyAlaGluPheLeuAlaGluGlyArgLys-----IleProGlySerAspIleAsp 412
Db 1676 GCTGCCCAAGAAACAGCAAGAAAGCCAAACCAAGTGGTAAATTTGCCGAACACGACCACT 1735
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTyrLys 426
Db 1736 TTGGCGAACAAGAGAAA-----ACAGGTACTTTGGAAA 1768
```

RESULT 35

```
US-09-815-242-4584
; Sequence 4584, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4584
; LENGTH: 6258
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4584
```

```
Alignment Scores:
Pred. No.: 0.716 Length: 6258
Score: 100.00 Matches: 103
Percent Similarity: 35.05% Conservative: 81
Best Local Similarity: 19.62% Mismatches: 215
Query Match: 3.15% Indels: 126
DB: 10 Gaps: 25
```

US-09-857-669-2 (1-609) x US-09-815-242-4584 (1-6258)

```
Qy 80 GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsn--- 98
Db 2899 GAAATATATTTTAAATAAATAAGTGGATCAAAATTTAGACAAAGCTCCGTTTGAAACGG 2958
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGlu--- 117
Db 2959 TTGTCACAGTGCAAAATGCCAAGAGTGCCCTAAATGGTAACCATATTTAGAGCAAGCT 3018
Qy 118 LysAspGlyAlaTyrThrVal-----HisIleThrProGlyProArgThr 132
Db 3019 AAATCAATGCAACACTACTATAAACCGACTTCAACATTTAACAACCTGCACAAAAGAT 3078
Qy 133 LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
```


Qy	80	GluGluValLeuAspLysGluClnThrGlyPheLeuAlaGluGluAlaProaspAsn---	98
Db	9730	GAATAATTTAAATATAAAAATAGTGGATCAAAATTTAGACAAAGCTGCCTGTCAAACCGC	9789
Qy	99	VallysThrMetLeuArgSerLysLysTyrrPheSerSerLysValSerLeuThrGlu---	117
Db	9790	TTCACCAAGTGACAAATGCCAAGGTGCCCTTAATGTATACCATAATTTAGCCCAAGCT	9849
Qy	118	LysaspGlyAlaTyrrThrVal-----HisIleThrProGlyProArgThr	132
Db	9850	AAATCAAAATGCCAACACTACTATAAACGGACTTCAACATTTTAACAACATGCACAAAAAGAT	9909
Qy	133	LysIleAlaasnValGlylValAlaIleLeuGlyAspIleLeuSerSerGlyAsnLeuAla	152
Db	9910	AAATGTG-----	9915
Qy	153	GluTyrrTyrrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheaspGln	172
Db	9916	-----AAACAACAAGTGCACAACCAACAANAATTGTCAGGT-----GTA	9954
Qy	173	AspSerTrpGluasnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrrPro	192
Db	9955	GATACTGTTAAATCAAGTGCCAACACATTAATGGTGCT-----	9993
Qy	193	LeuAlalysLeuGlyAsnThrArgAlaAlaValasnProaspThrAlaThr-	209
Db	9994	-----ATGGGTGACGTTAAGAANAATAGCATTCGAAGATAACACAGCTACCAATAAATGCC	10044
Qy	210	-----ValasLeuasnValvaliaspSer	218
Db	10045	CAAAACTATCTTGATGCTACAGAAAGTAACAAAACAACTATACAATGCTGTGTGATAGT	10104
Qy	219	GlyArgProIleAlaPheGlyaspPheGluIleThrGlyThrGlnArgTyrrProGluGln	238
Db	10105	-----GCTAATGGTGTCTATTAAATGCAACAGCAATCCAAAATATGGATGCTAAT	10152
Qy	239	IleValSerGlyLeuAlaAargPheGlnProGlyThrProTyrraspLeuaspleuLeuLeu	258
Db	10153	GCAATTAACCAAAATCGCTACCAAGTGACATCAACAGAAAATGCAATTAGATGGTACACAT	10212
Qy	259	AspPheGlnGlnAlaLeuGlu---GlnAsnGlyHistrSerGlyAlaSerValGlnAla	277
Db	10213	AATTTAAGCAAGCGAACAACAACAGCAACAATGCCATCGATGGTGCTACT-----	10263
Qy	278	AspPheaspArgLeuGlnGlyaspArgValProVallysValSerValThrGluVallys	297
Db	10264	AACTTAATAAAGCACACAANAAGAT-----CGCTTAAAGCAACAAGTTACAAGTCCGCAA	10317
Qy	298	ArghIsIlysLeuGluThrGlylleArgLeuaspSerGluTyrrGlyLeuGlylGyllysile	317
Db	10318	CGTGTTCGAATGTAAACAGTAGTATCCACAACACTGCA-----	10353
Qy	318	AlatyrAspyrrTyrrAsnLeuPheAsnLysglytyrIledgyservalValtrpAspmet	337
Db	10354	-----AATGAACCTTAATACAGCT---ATGGGTCAATTTACACATCGGTATT	10395
Qy	338	AsplysTyrrGluThrThrLeuAlaAleGlyIleSerGlnProArgAsnTyrrArglyasn	357
Db	10396	GATGATGAAAACGCAACA-----AAACAAMACTCAAAAATATCGTGACGCT	10440
Qy	358	TyrrprThrSerAsnValSerTyrr-----Asn	366
Db	10441	GAACAAGATAAGANAACCTGCTTATGATCAAGCTGTAGCTGCGCAAGCAATTTTTAAAT	10500
Qy	367	ArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlylGlyIleTrpTyrrVal---	385
Db	10501	AAACAACACTGGTCCAAATTCAGATAAAGCAGCACTTGCCCGTGCATTACACAAGTRACA	10560
Qy	386	-----ArgaspArgAlaGlylleaspAlaArgLeuGlyAlaGluPheLeualagLucly	403
Db	10561	AGTACGAAGATGATGTAATGGTGTGATGCGNAACT---GCAGAAGCGAAGCGGCAGCT	10617

QY 404 ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
 Db 10618 AAAAAAACTTTAGGCTACTTTAAACCATTAATGCAACGCTACTGCTTGAAGGT 10677
 QY 424 SerTPlLysArgGlnLeu-----LeuAsnAsnValLeuHisProGluAsnGly 439
 Db 10678 CAATCAATCAAGCCGACGACTGTCATGGCGTTAATACTGTG-----AAACAATAATGCC 10731
 QY 440 HistYrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459
 Db 10732 AATACATTAGACGGCGCAATGAATAGCTTCAAGGTTCAATCAATGATAAAGATGCGACA 10791
 QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys----- 476
 Db 10792 TTAAGAAATCAA-----AATTATCTTGATCGACAGTGAATCAAAACGTAACGCATAT 10842
 QY 477 -----LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAla 491
 Db 10843 ACTCAAGCTGTCAACAGCAGCTCAAGGCATTTTAAATAAACAACCTGGTGGTTAACACATCT 10902
 QY 492 ArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer-----GlyGly 507
 Db 10903 AAAGCAGACGCTGATAATGCAATTAATACAGTTACAACAGCGAAAGCGCTTTAAATGGT 10962
 QY 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
 Db 10963 GCTGAAACTTTAAGAAATACGAAACTTCACAAACAAATACGATTT-----AATGGG--- 11013
 QY 528 ValLeuProGlu---ArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThr 546
 Db 11014 ---TTACCAAACTTAACACAATTAACAAAAGATAACTTGAAGCATCAAGTTGAACAACCG 11070
 QY 547 ArgThrLeuSerGly 551
 Db 11071 CAAAATGTAGCAGGT 11085
 RESULT 37
 US-09-782-378A-26
 ; Sequence 26, Application US/09782378A
 ; Patent No. US20020102731A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hearing, Patrick
 ; APPLICANT: Bahou, Wadie
 ; APPLICANT: Sandalon, Ziv
 ; APPLICANT: Gnatenko, Dmitri
 ; TITLE OF INVENTION: Adenoviral Vectors
 ; FILE REFERENCE: STONYB-04970
 ; CURRENT APPLICATION NUMBER: US/09/782.378A
 ; CURRENT FILING DATE: 2001-02-12
 ; PRIORITY APPLICATION NUMBER: 60/237,747
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 26
 ; LENGTH: 35100
 ; TYPE: DNA
 ; ORGANISM: Human adenovirus type 17
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (25184)..(25184)
 ; OTHER INFORMATION: The n at this position can be a, c, t, or g.
 US-09-782-378A-26

```
Qy 156 ArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrp 175
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11796 AGCGCGCGCTGCAAGC-----CAGCGTCCGCT-----CTAGCATCGGACGACTGG 11843

Qy 176 GluAsnSerLysThrSerValLeuGlyAlaValThrArg-----LysGlyTrpProLeu 193
   ||| ||| : : : : : ||||| : : : : :
Db 11844 GAGCGTGGGTGCAACGATCATGCCCTGACGACCCCAACCCCGACTCTTTAGACAA 11903

Qy 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
   ||| ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 11904 CAGCCGCGAGCCACAGACTCTCGGCCATT-----CTGGAG 11939

Qy 214 ValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGln 233
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11940 GCGGTGGTCCCTCT-----CGGACCAACCCACGAC 11972

Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
   ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 11973 GAGAAGTGTGGCGATCGTAACGCGCTGGCGGAGAACAAAGGCCATCCGTCGCCGACGAG 12032

Qy 254 LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
   ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12033 CCGCGGTAGT-----TACACGCCCTGTGGAGCGGTAGGCGGTACACAGCAC 12086

Qy 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
   : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12087 AACGTGCAGTCCAACTCGACCGCTGTGTGACGGAC-----GTGCGGAGCCGCTG 12137

Qy 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
   : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12138 GCGAGCGCGAGCGGTTCAGAACAG-----GGCCTG 12170

Qy 314 GlyGlyLysIleAlaTyrAspTyrTrpAsnLeuPheAsnLysGlyTyrIleGlySerVal 333
   ||| ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12171 GCCTCGCTGGTGGC-----CTGAACGCTCTTCCTGGGACCCAGCCGCGGACGCTG 12221

Qy 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAla-----Gly 347
   ||| ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12222 CCGCGCGGCGAGGATGATTACACCAACTTTATCAGCGCGCTGCGCTGATGTCGCGAG 12281

Qy 348 IleSerGlnProArgAsnTyrArg--GlyAsnTyrTrpThrSerAsnValSerTyrAsn 366
   : : : ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12282 GTGCCCGCAGAGCGAGGTACAGTCGGCGCGGACTACTTTTCCAACTAGCAGACAG 12341

Qy 367 ArgSerThrThrGlnAsnLeuLysArgAlaPheSerGly-----GlyIleTrpTyr 384
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12342 GCCTGCAACCGTGAACCTG---AGCCAGGCTTTCAAGAACCTCGCGGGCTGTGGGC 12398

Qy 385 ValArg-----AspArgAlaGlyIleAspAlaArgLeu 395
   ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 12399 GTGCAGGCGCGGTGGCGACCGGTGCGAGCGGTGAGCAGCTTGTCTG 12443
```

RESULT 38

```
US-09-974-300-7800
; Sequence 7800, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7800
; LENGTH: 2085
; TYPE: DNA
```

```
; ORGANISM: Bacillus clausii
US-09-974-300-7800
```

Alignment Scores:

Pred. No.:	0.15	Length:	2085
Score:	99.50	Matches:	127
Percent Similarity:	33.80%	Conservative:	68
Best Local Similarity:	22.01%	Mismatches:	191
Query Match:	3.14%	Indels:	191
DB:	10	Gaps:	38

US-09-857-669-2 (1-609) x US-09-974-300-7800 (1-2085)

```
Qy 93 GluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLys 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 GAAGAGCGAGTGAGTGTTCATCCGTGG-----GAAACGGCTATATGAGTGGCG 422

Qy 113 ValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrProGlyProArgThr 132
   : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 423 TTGCGAAAGCCGAG---GATGGT-----GTAACCCCTGCAGGAGAGCG 464

Qy 133 LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
   : : : : : : : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 465 TACAGTACCATGCCAATGACATGATGTCGCGGATGTCATGCGGATGCGCATGATGAG 524

Qy 153 GluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGln 172
   : : : : : : : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 525 TTTTGTGAAA-----TATATTGATGCTATACATTGACTGCG 536

Qy 173 AspSerTrpGlu-----AsnSerLysThrSerValLeuGlyAlaValThrArgLysGly 190
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 -----TGGGAACCGCACAAATCTTAAGAC-----GTGTCACAAAAAGGT 575

Qy 191 TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 576 TAC-----ACCGAAACACC-----TATATTGATGCTATACATTGACTGCG 617

Qy 211 AspLeuAsnValValLysSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
   ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 618 GATTGTCTATACCGGATAGACTTGTGTAANAACATCCGTCCTCGGG-----562

Qy 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThr 250
   ||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 663 -----GCTCATACACGACGCTGCTGTT-----686

Qy 251 ProTyrAspLeuAsp-----LeuLeuLeuAspPheGlnGlnAlaLeuGluGln 266
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 687 ---TACGATTTGATGGCGACGCGCAAGCGGAACTGATGTTTAAACTGCA-----734

Qy 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---285
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 735 -----CCAGGGACAAGGTGATCACA-----TTCACTCGGATGCG 770

Qy 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
   : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 771 AAAGTCGAATCGAAAACGTATATTACG---ATGCCCGAGGAAGACGCTGGAAGCAGGTTAC 827

Qy 306 ArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
   : : : : : : : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 828 AGCCATGAAGACGATTACCGGATGAGT-----GCCGATGATTATTATATCCCATG 878

Qy 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLys-----339
   : : : : : : : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 879 GTCGAGATGTTTATTCGGT-----TGGCATGACCATGAAGAACTAAAAACCGCAAT 929

Qy 340 TyrGluThrThrLeu-----AlaAlaGlyIleSerGlnProArgAsnTyr-----354
   : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 930 TGGCCAGAGACGCTTGAAGAGCGCTTCGGGATCGAACCCTCTATGACTATCTTTAAGC 989

Qy 355 ArgGly-----AsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
   ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 990 CCGCGAGATGCGGAGCGCTTAGCCGATTATTTATTCGATGAATACGCCCTTCGAGAACT 1049
```

US-09-957-669-2 (1-609) x US-09-938-842A-1039 (1-3078)

Qy	109	PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro	128
Db	121	TTTCACCTCTACAATCAATCAAAAGCGACGCTGCT	165
Qy	129	GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp	148
Db	166	GTTCACAGCTCCT	213
Qy	149	GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer	168
Db	214	GGACACGACGTGTGTATCTAGAGGAGTTCACAAGAGCTTGGGAGCTGATCCACACAGT	273
Qy	169	AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArg	188
Db	274	-----GTTGATGAGTCGTGGGATAAATTTTGGAAATTTTCTGGGTGAG-----	318
Qy	189	LysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla	208
Db	319	-----GCCTTACATCGCGTGGTATCTCGGGGCAAAACCATTT	354
Qy	209	ThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGlu	228
Db	355	CAAGAAAGCATGCGTTTGGTTG-----TTGCTAGTTAGAGCTTACCAG	396
Qy	229	IleThrGly-----ThrGlnArgTyrPro	236
Db	397	GTTAATCGCCACATGAAGGCCAGCTTGATCCTTTAGTGCTACAGAAAGAGAGATCCCA	456
Qy	237	GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu	256
Db	457	GAGGATCTCAGCCAGCTCTTTATGGTTC-----ACTGAGGCTGATCTTGATCGG	507
Qy	257	LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln	276
Db	508	GAATTC--TTTCTGGTGTATGGAGGATGTCGGGTTTTCTCTCT-----	549
Qy	277	AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal	296
Db	550	-----GAGAACCGCCGGTT-----CAACA	570
Qy	297	LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLys	316
Db	571	CTGAGGTCGATACTGTCG-----AGGCTTGACCAAGCTTAC-----TGTGGGACT	615
Qy	317	IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp	336
Db	616	ATAGGGTATCAGTACATGCACATTGCTGATAGGGATAAATGTAAAC-----TGCTTG	666
Qy	337	MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly	356
Db	667	AGACACAAGATCGAGACC-----CCAACTCCTCGACAGTAC-----	702
Qy	357	AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg	376

Qy	277	AlaaspPheaspArgLeuGlnGlyaspArgValproVallysValSerValThrGluVal	296
Db	550	-----GAGAACGCCCGGT-----CAAACA	570
Qy	297	LysargHisLysLeuGluThrGlyIleArgLeuaspSerGluThrGlyLeuGlyGlyLys	316
Db	571	CTGAGGTGCGATACTGTCG-----AGCGTTGACGACGCTTAC-----TCGTGGACT	615
Qy	317	IleAlaIleAspTyrThrAsnLeuPheAsnLysGlyTyrIleGlySerValValTpaasp	336
Db	616	ATAGGGTATCAGTACATGCACATTGCTGATAGGATAAATGTAAC-----TGCTTG	666
Qy	337	MetaspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly	356
Db	667	AGACAGAAGATCAGACCC-----CCAACTCCTCGACAGTAC-----	702
Qy	357	AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg	376

```

Db 703 AATAGTGAGCGTGGATGGTATTATGATAGG---CTTACCTGGAGCACACAGTTTGAG 759
Qy 377 AlapheSerGlyGlyIleTTPtYrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 760 AATTCTTGGCTACTAGTGACACCGCTAAAGGTTTGGACTGGAA-----GGT 810
Qy 397 AlaGluPheLeuAlaGluGlyArgLys-----IleProGly 408
Db 811 GCTGAATCTTTGATTCCTGGCATGAAGAGATGTTCCATAGGTGCGAGATCTCGGGTA 870
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGln 428
Db 871 GAGAACATAGTATTCTCGGTATGCCCATAGGCGTGGTACTT-----AAT 912
Qy 429 LeuLeuAsnValLeuHis----- 435
Db 913 GTTTGGGTAAATGTTGTGAGAAACCTCTAGGCCAAATATTCACGCGAGTTTAGCGGTGGT 972
Qy 436 -----ProGluAsnGlyHisTrpLeu---AspGlyLysIleGlyThrTrpLeu 450
Db 973 ACTAGGCCAGTAGATGAAGTTGGCTTTACACCGAAGAGGTGATGTGAATACCACTTG 1032
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTrpPhePhe 470
Db 1033 GGTACATCTTATGATGCTCCAACTAGAGGAGGCAACATCTCCACTTGTCTTGGTAGCA 1092
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla-----Gly 487
Db 1093 AATCCCACTAGCTGGGAAGCAGTAGACTCTGTTGTGATAGGTAAACACGAGCAACAA 1152
Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
Db 1153 TATTACAGGAAGACGAGACAGACAGAACAGATGGTATTTGATCCATCGCGGATGGT 1212
Qy 508 AlaSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGly---ProAsn--- 525
Db 1213 AGCTTTCCGCGACAAGCAGGTGCTATGAAACTCTCCATCTTAGTGCACTTCCCTAACTAC 1272
Qy 526 -----GlySerValLeuProGluArgAlaLeuLeuValGly----- 537
Db 1273 TGTACCGGTGGACAGTG-----CACATTGTGTGTAATAATCAAGTGGCTTTC 1320
Qy 538 -----SerLeuGluTyrGlnLeuProPheThrArgThrLeu 549
Db 1321 ACAACCATCCCGGGAAGGAGTCTTCACAGTATTGCACTGATGTTGCCAAGCGCTTG 1380
Qy 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
Db 1381 AGCGCCCAATTTTCCATGTCAATGCAGATGCATTTGAA-----GCAGTA 1425
Qy 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
Db 1426 GTGATGCTTGTGAGCTTGCTGCTGAGTGG-----CQCCAGACGTTCCATTCGTAT 1476
Qy 590 IleAla-----TyrGlyHisSerAsp----- 596
Db 1477 GTTGTGTTGATTAGTATGCTACCGTGGCTTTGGGATACAGAGATAGAGAACCGTCA 1536
Qy 597 -----LysLysIleArgTrpHisIleSer 604
Db 1537 TTCACACAACCAAAATGTACAAGGTGATAGCGAGTCATCCCTCG 1581
```

RESULT 40

```

US-09-897-056-6
; Sequence 6, Application US/09897056
; Patent No. US20020100074A1
; GENERAL INFORMATION:
; APPLICANT: JOE, YUJI
; APPLICANT: MIWA, TERSUYA
; APPLICANT: KIDA, TAKAO
; APPLICANT: IGARASHI, DAISUKE
; APPLICANT: OHSUMI, CHIEKO
```

```

; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSFORMED PLANT HAVING INCREASED GLUTA
; FILE REFERENCE: 210425US0
; CURRENT APPLICATION NUMBER: US/09/897,056
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: JP 2000-215279
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(3152)
; OTHER INFORMATION:
US-09-897-056-6

Alignment Scores:
Pred. No.: 0.321 Length: 3412
Score: 99.50 Matches: 113
Percent Similarity: 31.83% Conservative: 70
Best Local Similarity: 19.65% Mismatches: 225
Query Match: 3.14% Indels: 167
DB: 10 Gaps: 30

US-09-857-669-2 (1-609) x US-09-897-056-6 (1-3412)
Qy 109 PheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 198 TTTCACTCTCAATACTCAAACTCAAGGAGGAGCTCTCT-----GCACCT 242
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 243 GTTCCACGCTCT-----GTCCCACTTTCTAAGCTAAGCTAGCTCTTTAGAT 290
Qy 149 GlyAsnLeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
Db 291 GGAAACAGCAGTGTATCTAGAGGAGTTCAAGAGCTTGGGAGGCTGATCCCAACAGT 350
Qy 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArg 188
Db 351 -----GTTGATGAGTCTGGGATACTTTTATAGGAATTTTGGTGGTCAG----- 395
Qy 189 LysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla 208
Db 396 -----GCTTCTACATCGCTGCTGATCTCGGGGCAACCACTT 431
Qy 209 ThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGlu 228
Db 432 CAAGAAAGCATGCTTTGTTG-----TTGCTAGTTAGAGCTTACCAG 473
Qy 229 IleThrGly-----ThrGlnArgTyrPro 236
Db 474 GTTAATGGCCACATGAAGCCCAAGCTTTAGTCTCTAGAGAGAGAGATTTCCA 533
Qy 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTrpAspLeu 256
Db 534 GAGGATCTCAGCCGCGCTTTATGGGTTT-----ACTGAGGCTGATCTTGTATCGG 584
Qy 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTrpSerGlyAlaSerValGln 276
Db 585 GAATTC---TTCTGGGTGATGAGGAGTTCGGGTTTTCTCTCT----- 626
Qy 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
Db 627 -----GAGAACCGCCCGGTT-----CAACA 647
Qy 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLys 316
Db 648 CTGAGGTCGATACCTGCTCG-----AGGCTTGACCAAGCTTAC-----TGTGGGACT 692
```

Search completed: November 9, 2002, 03:59:22
Job time : 497 secs

Qy	317	IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp	336
Db	693	ATAGGTAAGTACATGCACATATGCTGATAGGATAAAATGTAAC	743
Qy	337	MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly	356
Db	744	AGACACAGATCGAGACC	779
Qy	357	AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArg	376
Db	780	AATAGTAGGCTCGATGGTATTATTTATGATAGG	836
Qy	377	AlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly	396
Db	837	AATTTCCTGGCTACTAAGTCGACCCAGCGGTAAAGGTTTGGACTGGAA	887
Qy	397	AlaGluPheLeuAlaGluGlyArgLys	408
Db	888	GCTCAATCTTTGATTCCTGGCATGAAGAGATCTTCGATAGTCTGCAGATCTCGGGTA	947
Qy	409	SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGln	428
Db	948	GAGAACATAGATTATCGGTATGCCCATAGGGTTCGACTT	989
Qy	429	LeuLeuAsnAsnValLeuHis	435
Db	990	GTTTGGGTAAATGTTGTAGAAAACCTCTACGCCAAATATTCAGCGAGTTTAGCGGTGGT	1049
Qy	436	ProGluAsnGlyHisTyrLeu	450
Db	1050	ACTAGGCCAGTAGATGAAGTTGGGCTTTACACCGGAACAGGTGATGTCAAATACCACTTG	1109
Qy	451	GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe	470
Db	1110	GGTACATCTTATGATCGTCCACTAGAGAGGCCAAACATCTCCACTGTCTCTTGGTAGCA	1169
Qy	471	ThrProGluAsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAla	487
Db	1170	AATCCCACTCACTTGGNACCACTAGATCTCTGTGTGATAGTTAAACCCAGAGCCAAACAA	1229
Qy	488	TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly	507
Db	1230	TATTACAGAAACAGCAGACAGAACAAAGACATGGGTATTTGATCTCATGGGGATGGT	1289
Qy	508	AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGly	525
Db	1290	AGCTTTCGGGACAGAGGTGGTGTATGAACTCTCCATCTTAGTCACCTTCCTTAACATAC	1349
Qy	526	GlySerValLeuProGluArgAlaLeuValGly	537
Db	1350	TGTACCGGTGAAACAGTG	1397
Qy	538	SerLeuGluTyrGlnLeuProPheThrArgThrLeu	549
Db	1398	ACAAACGNTCCAGGAAGAGGTCTTCACAGTATGTGCATGTGTTGCAAAAGGCTTTG	1457
Qy	550	SerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLysLeu	569
Db	1458	AGGCCCAATTTTCCATCTCAATCCAGATGACATTGAA	1502
Qy	570	LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp	589
Db	1503	GTGCATGCTGTGAGCTTGCTCTCAGTGG	1553
Qy	590	IleAla	596
Db	1554	GTTCCTGTTGATTAGTATGCTACCGTGCCTTTGGCCATACAGAGATAGACGAACCGTCA	1613
Qy	597	LysLysIleArgTrpHisIleSer	604
Db	1614	TTCACACACCAAAATGTACAGAGGTGATAGCGACTGATCCTCTCC	1658

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:14:59 ; Search time 32 Seconds
(without alignments)
3921.334 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rviri:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3153	99.4	615	16	Q9JWM5	Q9jwm5 neisseria m
2	3136	98.8	635	16	Q9JXB7	Q9jxb7 neisseria m
3	596	18.8	589	16	Q8XWT6	Q8xwt6 ralstonia s
4	449	14.2	579	16	Q910U1	Q910u1 pseudomonas
5	368.5	11.6	578	16	Q8ZB95	Q8zb95 yersinia pe
6	355.5	11.2	577	16	Q8XEL5	Q8xel5 salmonella
7	348	11.0	582	16	Q9KP31	Q9kp31 vibrio chol
8	340	10.7	618	16	Q92LN9	Q92ln9 rhizobium m
9	336	10.6	586	16	Q9CK26	Q9ck26 pasteurella
10	322.5	10.5	765	16	Q8XZ13	Q8xz13 ralstonia s
11	323.5	10.2	641	16	Q8UC83	Q8uc83 agrobacteri
12	306.5	9.7	617	16	Q9PD27	Q9pd27 xylella fas
13	300	9.5	738	2	Q9X3V8	Q9x3v8 zymomonas m
14	292	9.2	593	2	Q9EY32	Q9ey32 xanthomonas
15	290.5	9.2	623	16	Q8YE13	Q8ye13 bruceella me
16	279	8.8	833	16	Q8YUR6	Q8yur6 anabaena sp

17	272.5	8.6	628	16	Q9A7W6	Q9a7w6 caulobacter
18	260.5	8.2	792	2	P95359	P95359 neisseria g
19	253	8.0	797	16	Q9K1H0	Q9k1h0 neisseria m
20	253	8.0	797	16	Q9JX31	Q9jx31 neisseria m
21	251	7.9	797	2	O30912	O30912 neisseria m
22	246	7.8	778	16	O67326	O67326 aquifex ae
23	246	7.8	795	16	O82H58	O82h58 yersinia pe
24	240.5	7.6	617	16	Q98K30	Q98k30 rhizobium l
25	228.5	7.2	804	16	Q8ZRP0	Q8zrp0 salmonella
26	227	7.2	797	2	Q9S341	Q9s341 photorhabdu
27	227	7.2	803	16	Q9KRW0	Q9kpw0 vibrio chol
28	226	7.1	803	16	Q8Z9A3	Q8z9a3 salmonella
29	219	6.9	792	2	O32625	O32625 haemophilus
30	217	6.8	785	2	Q9R2E3	Q9r2e3 escherichia
31	215.5	6.8	797	16	Q9HXY4	Q9hxy4 pseudomonas
32	211	6.6	676	16	Q8YMP0	Q8ymp0 anabaena sp
33	210.5	6.6	769	16	Q9A711	Q9a711 caulobacter
34	206	6.5	739	16	Q9PI28	Q9pi28 campylobact
35	202.5	6.4	861	16	P73472	P73472 synchocyst
36	199	6.3	792	16	Q9PKF3	Q9pkf3 chlamydia m
37	197.5	6.2	784	16	Q9PEI2	Q9pei2 xylella fas
38	194.5	6.1	781	16	Q8YHH0	Q8yhh0 bruceella me
39	193.5	6.1	793	2	Q93PM2	Q93pm2 haemophilus
40	189.5	6.0	776	16	Q92Q48	Q92q48 rhizobium m
41	185	5.8	774	16	Q8UFL6	Q8uf16 agrobacteri
42	179	5.6	491	16	Q8ZOL6	Q8z016 anabaena sp
43	179	5.6	789	2	O51930	O51930 pasteurella
44	178.5	5.6	794	16	Q98MC3	Q98mc3 rhizobium l
45	177	5.6	678	16	Q8RHT4	Q8rht4 fusobacteri

ALIGNMENTS

RESULT 1

Q9JWM5	PRELIMINARY;	PRT;	615 AA.
AC	Q9JWM5;		
DT	01-OCT-2000 (TREMREL. 15, Created)		
DT	01-OCT-2000 (TREMREL. 15, Last sequence update)		
DT	01-DEC-2001 (TREMREL. 19, Last annotation update)		
DE	Putative outer membrane protein.		
GN	NMA0296.		
OS	Neisseria meningitidis (serogroup A).		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=65699;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;		
RX	MEDLINE=20222556; PubMed=10761919;		
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,		
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,		
RA	Davies R.M., Davis P., Deakin K., Feltwell T., Hamlin N., Holroyd S.,		
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,		
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,		
RA	Whitehead S., Spratt B.G., Barrall B.G.;		
RT	*Complete DNA sequence of a serogroup A strain of Neisseria		
RT	meningitidis 22491.*;		
RL	Nature 404:502-506(2000).		
DR	EMBL; AL162752; CAB83602.1; -.		
KW	Complete proteome.		
SQ	SEQUENCE 615 AA; 67775 MW; B9A64607638E1CCB CRC64;		
Query Match 99.4%; Score 3153; DB 16; Length 615;			
Best Local Similarity 99.3%; Pred. No. 2.7e-192;			
Matches 605; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			
Oy	1	MMIKPTALLPALFFPHAYAPADLSNKAAAGFALFKNKSPDTSVKLPKFPVRIQTQ	60
Db	7	MMIKPTALLPALFFPHAYAPADLSNKAAAGFALFKNKSPDTSVKLPKFPVRIQTQ	66
Oy	61	DSEIKMDVEEHLPLITOOQEEVLDKEOTGFLAEAPDNVKTMLRSGYFSSKVSITKDG	120

Db	67	DSEIKDMVEEHLPLITQOOEEVLDKEQTCFLAEAPDNVKTMLRSKGYFSSKVSUTEKDG	126
Qy	121	AYTVHITPGPRTKIANVGVAIILGDILSDGNLAEEYRNALENMQQPVGSDFDODSWSKNT	180
Db	127	MMIKETALLPALFFPHAYAPAADLSENKAAGFALFNKNSPDTESVKLKPKFPVIDTQ	86
Qy	181	SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDGSRPIAFGDFEITGTQRYPEQIV	240
Db	187	SVLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVVDGSRPIAFGDFEITGTQRYPEQIV	246
Qy	241	SGLARFQPGTPTDLDLLDFQALBQNGHYSGASVQADFDRLQGDVPRVKVSVTEVKRHK	300
Db	247	SGLARFQPGTPTDLDLLDFQALBQNGHYSGASVQADFDRLQGDVPRVKVSVTEVKRHK	306
Qy	301	LETGIRLSEYGLGCKIAYDYNNLKNKGYIGSVVWMDKYEITTLAAGISOPRNYRGNTWT	360
Db	307	LETGIRLSEYGLGCKIAYDYNNLKNKGYIGSVVWMDKYEITTLAAGISOPRNYRGNTWT	366
Qy	361	SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM	420
Db	367	SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM	426
Qy	421	LTASWKROLLNNVLPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKLGTF	480
Db	427	LTASWKROLLNNVLPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKLGTF	486
Qy	481	IIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSITGLAGPNSVLPERALLVGSLE	540
Db	487	IIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSITGLAGPNSVLPERALLVGSLE	546
Qy	541	YOLPFTRTLGSAGVFDHMDGDAANFRMKLKHGSGGLGVRFWSPLAFSFDIAYGHSDDKIR	600
Db	547	YOLPFTRTLGSAGVFDHMDGDAANFRMKLKHGSGGLGVRFWSPLAFSFDIAYGHSDDKIR	606
Qy	601	WHISLGTFR 609	
Db	607	WHISLGTFR 615	
RESULT 2			
QJXB7	QJXB7	PRELIMINARY:	PRT: 635 AA.
AC	QJXB7		
DT	01-OCT-2000 (TremBLrel. 15, Created)		
DT	01-OCT-2000 (TremBLrel. 15, Last sequence update)		
DT	01-MAR-2002 (TremBLrel. 20, Last annotation update)		
DE	Hypothetical protein NMB2134.		
GN	NMB2134.		
OS	Neisseria meningitidis (serogroup B).		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=491;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC58 / SEROGROUP B;		
RX	MEDLINE=20175755; PubMed=10710307;		
RA	Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,		
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,		
RA	Nelson W.A., Gwinn M.L., DeBoy R., Peterson J.O., Hickey E.K.,		
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,		
RA	Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,		
RA	Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,		
RA	Gill J., Scarlato V., Maignani V., Piazza M., Grandi G., Sun L.,		
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;		
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain		
RT	MC58."		
RL	Science 287:1809-1815(2000).		
DR	EMBL; AE002561; AAF42442.1; -.		
DR	TIGR; NMB2134; -.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 635 AA: 69801 MW; 2B371E4A7A9DDA98 CRC64;		
Query Match		98.8%;	Score 3136; DB 16; Length 635;
Best Local Similarity		98.7%;	Pred. No. 3.4e-191;

Matches	601;	Conservative	3;	Mismatches	5;	Indels	0;	Gaps	0;
QY	1	MMIKPTALLPALFFPHAYAPAADLSENKAAGFALFNKNSPDTESVKLKPKFPVRIDTQ	60						
DB	27	MMIKPTALLPALFFPHAYAPAADLSENKAAGFALFNKNSPDTESVKLKPKFPVLIDTQ	86						
QY	61	DSEIKDMVEEHLPLITQOOEEVLDKEQTGFLEAPDNVKTMLRSKGYFSSKVSLETKDG	120						
DB	87	DSEIKDMVEEHLPLITQOOEEVLDKEQTGFLEAPDNVKTMLRSKGYFSSKVSLETKDG	146						
QY	121	AYTVHITPGPRTKIANVGVAIILGDILSDGNLAEEYRNALENMQPVGSDFDQDSWENSKT	180						
DB	147	AYTVHITPGPRTKIANVGVAIILGDILSDGNLAEEYRNALENMQPVGSDFDQDSWENSKT	206						
QY	181	SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDGSRPIAFGDFEITGTORYPEQIV	240						
DB	207	SVLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVVDGSRPIAFGDFEITGTORYPEQIV	266						
QY	241	SGLARFQPGTPTDLDLLDFQALBQNGHYSGASVQADFDRLQGDVPRVKVSVTEVKRHK	300						
DB	267	SGLARFQPGMPYDLDLLDFQALBQNGHYSGASVQADFDRLQGDVPRVKVSVTEVKRHK	326						
QY	301	LETGIRLSEYGLGCKIAYDYNNLKNKGYIGSVVWMDKYETTTLAAGISQPRNYRGNYWT	360						
DB	327	LETGIRLSEYGLGCKIAYDYNNLKNKGYIGSVVWMDKYETTTLAAGISQPRNYRGNYWT	386						
QY	361	SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM	420						
DB	387	SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAELAEGRKIPGSAVDLGNSHATM	446						
QY	421	LTASWKROLLNNVLPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKLGTF	480						
DB	447	LTASWKROLLNNVLPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKLGTF	506						
QY	481	IIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSITGLAGPNSVLPERALLVGSLE	540						
DB	507	IIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSITGLAGPNSVLPERALLVGSLE	566						
QY	541	YQLPFTRTLGSAGVFDHMDGDAANFRMKLKHGSGGLGVRFWSPLAFSFDIAYGHSDDKIR	600						
DB	567	YQLPFTRTLGSAGVFDHMDGDAANFRMKLKHGSGGLGVRFWSPLAFSFDIAYGHSDDKIR	626						
QY	601	WHISLGTFR 609							
DB	627	WHISLGTFR 635							
RESULT 3									
QBXWT6									
AC	QBXWT6	PRELIMINARY;		PRT;	589	AA.			
ID	QBXWT6;								
DT	01-MAR-2002 (TremBLrel. 20, Created)								
DT	01-MAR-2002 (TremBLrel. 20, Last sequence update)								
DT	01-MAR-2002 (TremBLrel. 20, Last annotation update)								
DE	Putative lipoprotein.								
GN	RSC2384 OR RS02766.								
OS	Ralstonia solanacearum (Pseudomonas solanacearum).								
OC	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;								
OC	Ralstonia.								
OX	NCBI_TaxID=305;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=GM11000;								
RX	MEDLINE=21681879; PubMed=11823852;								
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangelot S.,								
RA	Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,								
RA	Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,								
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,								
RA	Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,								
RA	Weissenbach J., Boucher C.A.;								
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum.;"								
RL	Nature 415:497-502(2002).								
DR	EMBL; AL646069; CAD16091.1; -.								


```
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Kariyasev A.V.,
RA Leather S., Moule J., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414157; CAC92753.1; -.
DR InterPro; IPR000566; Lipocln_cytfabp.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 578 AA; 64405 MW; 69AD3780BFDC20F1 CRC64;

Query Match 11.6%; Score 368.5; DB 16; Length 578;
Best Local Similarity 22.8%; Pred. No. 3.4e-15;
Matches 153; Conservative 93; Mismatches 264; Indels 161; Gaps 21;

QY 1 MMTKPTALLLPALFFPHAYAP-----AADSENKAAGFALFNKNSPDFTESVKLKP 51
DB 1 MLRYPLICFVCLLATPIAYAAHVRVQLQVEGLSGDLERNVRLSTI-----GTDEVTADG 55
QY 52 KPVRIIDTODSEIKDMVEEHLPLITQQOEVLDKQETGFLAEEAPDNVKTMLRSKGYFSS 111
DB 56 RFRSRVD-----RAIROGLRALGYDP 77
QY 112 KVSLETKD-----GAYTVHTPGPKTKIANVGVAITLGDILSDGNLAEEYRNALENWQOP 165
DB 78 TITFELQNRPAARSVLIAKVVPCEPVLIAGVDIVLQGGAKTDPD---YQALVRRDTPK 133
QY 166 VGSDFDQDSEWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAG 225
DB 134 IGSVLNRHGFQDNFTSSLTGLALRGYFDANMIKSQGLVAAQLHEAFWDIDFDSGQRYRF 193
QY 226 DFEITGTQRYPEQISGLARFQPGTGYDLDLLDFOALEQNGHYSGASVQADF-DRLQG 284
DB 194 KVFQCSQ-IREDYLNQLNVPFHEGEFTYSDLAELNRRLAATNWFNSVVPDFQDAKES 252
QY 285 DRVVPKVSVEVKKHKLKLETGIRLDSYGLGKTIAYDYNNFNKGYIGSVVWMDKRYETTL 344
DB 253 KILPLDAVTP-----RTENTVELGGGVATD-----VG-----PRL 283
QY 345 AAGISQPNRYGNWYTSNYSNRSTQN-----EKRAFSGGIWYVRDRAGIDARLGAEPFL 400
DB 284 TASWRKP-----WNNSFGHSITTTTALSAPQETLDFSYRIPLLRN-----PLBQYYL 330
QY 401 AEG--RKIPGSDIDLGNSHATM-----LTASMRQLLNVLNHPENGHYLDKIGKTTL 450
DB 331 IQGQPR--TDLNDNTSDTTLNARFWDLSGQRAI---NLRNSLDHFTQGRVTDTT 384
QY 451 GTFLSTALIRTSARAGY-----FTPTEN---KKLG--- 478
DB 385 MLLYPGVSNRTRQRGAMPVWGDQSRYSDVSDTTWGSVDVDFGIFQAQNVWIRTLEKN 444
QY 479 TFIIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGNGSVLPERALLVGS 538
DB 445 RFYARNVGWGIETNFRDPPSLRFFAGGDRSIRGYKFRNISPRSDSGKLTGASKLATS 504
QY 539 LEYQLPFTRTLSCAFVDMGDAANFKRMKHKHSGSLGVKRWFPPLAFPSFDIAYGHSDKK 598
DB 505 LEYQYNTGRWAGVFDVSGEAVNNFSKSDLTGTCAGVGVWRWASVPVGIKLDIAAPICDNE 564
QY 599 ---IRWHISLG 606
DB 565 THGVQFVIGL 575

RESULT 6
Q8XELS PRELIMINARY; PRT; 577 AA.
AC Q8XELS;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
```

```
DE Putative outer membrane protein (Putative exported protein).
GN YTFM OR STM4409 OR STY4768.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RX [1]
RC SEQUENCE FROM N.A.
RC SPECIES-S typhimurium; STRAIN=LT2 / SCSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RX [2]
RC SEQUENCE FROM N.A.
RC SPECIES-S typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Hoiden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE008906; AAL23228.1; -.
DR EMBL; AL627283; CAD06889.1; -.
DR InterPro; IPR000184; Bac_surfag_D15.
DR InterPro; IPR000566; Lipocln_cytfabp.
DR Pfam; PF01103; Bac_surface_Ag_1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 577 AA; 64765 MW; FBA4A190173BF18C CRC64;

Query Match 11.2%; Score 355.5; DB 16; Length 577;
Best Local Similarity 22.3%; Pred. No. 2.2e-14;
Matches 134; Conservative 99; Mismatches 267; Indels 101; Gaps 17;

QY 55 VRIDTODSEIKDMVEEHLPLITQQOEVLDKQETGFLAEEAPDNVKTMLRSKGYFSSKVS 114
DB 26 LKVEGLSGELEKNVRAQLSTI--QSDEVTDDR---FRARVDDAIRGLKALGYEPTIK 80
QY 115 L-----TEKDGAYTVHTPGPKTKIANVGVAITLGDILSDGNLAEEYRNALENWQOPVGS 168
DB 81 FDLLPPAKGRQVLIARVTPGGPVLIGCTEVLURGARTD---KDYL--ALLKTRPAIGT 135
QY 169 DFDQDSEWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAGDFE 228
DB 136 VLNQGDYDNFKKSLTSVSLRKGYFDSEFIKSQGLGALGRHQAQFWDIDYDSGERYRFGPVT 195
QY 229 ITGTQRYPGQVSGLARFQGPPTPYDLDLLDFOALEQNGHYSGASVQADFQRLQGRV- 287
DB 196 FEGSQ-IRDEYLNQLLPPFKEGDEYESKDLAELNRRLSATGWFSVVVAPEFEKSRKTKIL 254
QY 288 PVKVSVEVKRHKLKLETGIRLDSYGLGKIAYDYNNFNKGYIGSVVWMDKRYETTLAAG 347
DB 255 PLKGVSPRTENTTETGTCVGTSDVGRPVKASW-----KKPWNSY-----CHSLTSTS 303
QY 348 ISQPNRYGNWYTSNYSNRSTQNLEKRAFSGGIWYVRDRAGIDARLGAELAEQ--RK 405
DB 304 ISAPEQ-----VLDFSYKMLPLKN-----PLEQYVLVQGGFKR 336
QY 406 IPGSDID-----LGNSHATMLTASWKRQLLNVLNHPENGHYLDCKIGTTLGLTFLSALLI 460
```

```
Db 337 TDLNTEQDSTTLAVSRWYDMLSSGQRAI---NLRSFDFHFTQGNVTNTMLFPGVMIS 393
Qy 461 RTSARAG-----YFPTPENKKGCT-----PIIRGOAGY 488
Db 394 KTRSGGLMPTWGDQSQRYSVDYNTANGSDVDVFLQAOQVWIRLYDRHRFVVRANLGM 453
Qy 489 TVARNADVPGLMFRSGGASVRYELDSIGLAGPNSVLPERALLVGSLEYQLPFTTRT 548
Db 454 IETGDFDKVPPDLRFAGGDRSIRGYKYKISPKSDSDGNLKGASKIATGSLEYQVYNTGK 513
Qy 549 LSGAVFHDGDAANAFKMKLKHGSLGVRNFSPLAFSFDIAYCHSDKK----IKWHLSL 605
Db 514 WMGAVFVDSGEAVSDIRRSDFKTGTGVRWASPVGPVKLDFAVPVGDKDEHGLQFYGL 573
Qy 606 G 606
Db 574 G 574

RESULT 7
ID O9KP31 PRELIMINARY; PRT; 582 AA.
AC O9KP31.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein VC2548.
GN VC2548.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004323; AAF56689.1; -.
DR TIGR; VC2548; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR InterPro; IPR000566; Lipocin_cytfABP.
DR Pfam; PF01103; Bac_surfaceAg; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 582 AA; 65644 MW; E36946FB85F1C426 CRC64;

Query Match 11.0%; Score 348; DB 16; Length 582;
Best Local Similarity 23.1%; Pred. No. 6.8e-14;
Matches 136; Conservative 96; Mismatches 255; Indels 102; Gaps 19;

Qy 65 KDMVEHLPLITQO-----EEVLDEKQGFLEAEAPDNVKTMLRSKGFSSKVSLT- 116
Db 46 KDNVEYLSIAAQDYSTSLRFQSOLERSMT-----EA-----LNAIGYHPHSIDFTV 93

Qy 117 -EKDGAYTVHTTPGPRTKIANGVAILGDLSDGNLAEYRNALENWQPVGSDFDQDSW 175
Db 94 SEDNORLRAAVTGEVTRUSEVDIVIRGEAGDRFQRLIRKS-----GLURVDAPLNHSLY 149

Qy 176 ENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGEITGTORY 235
Db 150 DNLSKSGIRNALQKGYFNFGDFQASRLVPELQARVILHFDSGIRYLFEGATTVESQ-I 208

Qy 236 PRQIVSGLARFQPGTPTDLDLLDFQQALEONGHYSGASVQADFRL-OGDRVPKVSVT 294
Db 209 DENRVMSLRPFKQGEPLYLVSVQGEFNQNLSTNDFSSVFEVPEPDLSQLDEGRELPKVTLA 268
```

```
Qy 295 EVKRHKLETGIRLDSEYGLGKGIAYDYVYNLFNKGVIGSVVWMDMKYETTLAAGISQP--- 351
Db 269 POARNQLETGYSTDVGVRSGLKWKPPVNSQGHSSFSLSLPEOTITAGYKIPLED 328
Qy 352 ---RNYRGNYTNSVYRSTTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPG 408
Db 329 ALNEYRYIOYGMKHL--DKRDTESLESNLSLRHW----- 361
Qy 409 SDIDLGNSHATMLTASWKRQLLNWV---LHPENGHY-LDGKIGITGLT----- 452
Db 362 -OLD-GGWHRTVFI---RYLLENYRQGLQDDNSQFLLPGMTYTRTRSRNSGLLTWGDK 415
Qy 453 -----FLSSTALIRTSARAGYFFT-PENKKGITFIIRGQAGYTVARNADVPFG 500
Db 416 QTITLEYGDPAALLSETVRVLRLOTGSSWLRTYARNIRA---LVRVGGANLVDEFQOLSPS 472
Qy 501 LMFRRGGASSVRYELDSIGLAGPNSVLPERALLVGSLEYQLPFTRTLSCGAVFHDGMDA 560
Db 473 LRFAGGDNLRGKYKISIPQDASGALTGAKYIATSIETSYQYRLTGNWAAWAFWVDVGDA 532
Qy 561 AANFKRMKLGKSGLVGRWFSPLAPFSFDIAYG---HSDKKIRWHISLG 606
Db 533 FND--NPEWKKGVGTGIRWISPVGPRIRLDFAWGLDAPGDEFKIHFTLG 579

RESULT 8
ID O92LN9 PRELIMINARY; PRT; 618 AA.
AC O92LN9.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein R02996.
GN R02996 OR SMC03097.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; ALS91792; CAC47575.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; trna-synt I.
DR Pfam; PF01103; Bac_surfaceAg; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 618 AA; 65656 MW; E51276BB29B1FB78 CRC64;

Query Match 10.7%; Score 340; DB 16; Length 618;
Best Local Similarity 22.0%; Pred. No. 2.4e-13;
Matches 144; Conservative 102; Mismatches 266; Indels 144; Gaps 22;

Qy 19 AYAPADLSEKNAAGFALFKNKSPTESVKLKPFPVRI-----DTQDSEIKDMVEHL 72
Db 2 ASAVLGPVSVEQAHAFKIFGMRFESAEVEQVIDPVRYTLTFEPTDDEELREALENG 61

Qy 73 PLITQQOEVLDEKQETGFLAEAPDNVKTMLRSKGFSSKVS-----TEKDGY- 122
Db 62 OLVDQOEPEVSGDLGLAIKARDRDLRLAVLEKARYGTVSILVNGDIDSLPPDPAPP 121
```

```

QY 123 -----TWHITPGPRTKIANVGVAILGDLSDGNLAAYYRNALENMPOQVGSDFDQDSW 175
Db 122 DGQPPVVRVAVPGP-----AFTLGTVRLEGDAARL-----DPAAYDLKRGAR 164
QY 176 ENSK-----TSVLGAVTRKGYPLAKLGNTRAAYNPDTATVNLNVVDSGRPIAFGDFEI 229
Db 165 ADSTLIKAGEQVNDLKEQSRPLAKLABRSVVADHATSTVDVTIRADGGPVAVPGLNLT 224
QY 230 TGTQRYPEQIVSGLARPOCTPYDLDLLDFOALEQNGHYSGASVQADFDRLQGD-RYP 288
Db 225 SGARTVDPDFVKDYSLNHNRPYSPENIRKAAERLQNLVFSVTIN-RADGLAPDGTIP 283
QY 289 VKVSTEVKHKLETGIRLDSEYGLGKIAYDYNNLFNKGYIGSVVWMDKY-ETTLAAG 347
Db 284 MNIVQSEGKHRVFGGQVSTTDLGLQGWGHRNLFGRASLRIEGSDRLGETTVDVAG 343
QY 348 ISOPRNYRGNWTSNVSYNRSTQNLEKRAFSGGIWYVRDRA----- 389
Db 344 LD-----YSGAILFAKPGAGFPASTFTASVKAIVDP 375
QY 390 -----GIDARLCAEF-----LABGRKIPGSDID--LGNSHAT-MLTASWKROLLNN 432
Db 376 DAYSAKTVTAAGAAPELSPEOTFSVGAECVADVDADFSGNSYITAALPFYVVRDARD 435
QY 433 VLPENG-----HYLDGKIGTTLTFLSSTALINTSARAGYFFTPENKKLGTFTII 482
Db 436 KLNPTGEGYRALINAKPSYIEGK-----TTFSS-----FEASAGYVAFGTEKR---FVL 482
QY 483 RQO--AGYTVARDN-ADVPISGLMFRSGASSYRGYELDSIGLAGPNSVLPERALLVGS 539
Db 483 AKLGLAGVLVGGDELSDIPATRRFFLGGGSGRGYSYQEISPRDADDELTTGGRSYVSGSL 542
QY 540 EYQLPPTRLTSLGAVFIDMG-----DAAANFKRMKLGKHSGLGVWRFSPAPFSDIA 591
Db 543 EARIATVDTIGVVPFIDACTVSDSTAPDS--DIRAGAGIGLRYATPPGPIRLDFA 596

```

RESULT 9

```

Q9CK26 PRELIMINARY; PRT; 586 AA.
AC Q9CK26;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2002 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein PM1809.
GN PM1809.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006218; AAK03893.1; -.
DR InterPro; IPR001084; Bac_surfaq_d15.
DR InterPro; IPR001092; HLH_basac.
DR Pfam; PF01103; Bac_surface_Ag; 1.
DR PROSITE; P500038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 586 AA; 67106 MW; 00895174DED58283 CRC64;

```

Query Match 10.6%; Score 336; DB 16; Length 586;

Best Local Similarity 25.0%; Pred. No. 4e-13; Mismatches 292; Indels 46; Gaps 15;

Matches 142; Conservative 88; Mismatches 292; Indels 46; Gaps 15;

QY 60 QDSEIKDWEEHLPLITTOOQEEVLDKEQTFGLAEAPDNVKTMLRSKGYFSSKVSU-----115

Db 41 QNTOLNNVRIYVGM1--DKEADGDSERHKQLVREALDKA---LRAYGYQSEVEFQIES 95

```

QY 116 --TEKDAGYVHTPGPRTKIANVGVAILGDLSDGNLAAYYRNALENMPOQVGSDFDQD 173
Db 96 QKPPKKQDLIAHVKIGKPVKIADVDFVLQGEATQDPEFIATVKNIPEK-----GSLNHE 150
QY 174 SHENSKTSVLGAVTRKGYPLAKLGNTRAAYNPDTATVNLNVVDSGRPIAFGDFEITGTQ 233
Db 151 TYDNYKSHIQKIALSRGYFDGDFLVSRLEVRPSTQAAWRLDYDSGERYRFGVEVTFENAO 210
QY 234 RYPEQIVSGLARPOCTPYDLDLLDFOALEQNGHYSGASVQADFDRLQGD-RYPVVKSV 293
Db 211 -TREDYLRNMIENFGOPYLINDLSTLTNNYTSNNWFSVLMQPVLD-EHKIVNVDVLL 268
QY 294 TEVKRHKLETGIRLDSEYGLGKIAYDYNNLFNKGYIGSVVWMDKYETTLAAG-----I 348
Db 269 OPRKNSMEVGIGWASDVGPRLQGLGWTKPWINNRHSFRTNLYVSAPKQTLATYKMPLL 328
QY 349 SOPRNYRGNWTSNVSYNRSTQNLEKRAFSGGIWYVRDRAIDARLGAELAEGRKIKG 408
Db 329 KNPLNYEYSAGLENENKNDTESFASL--SAIRYWNHEAGWQHSGLURVRYD----- 380
QY 409 SDIDLGNSHATWL---TASWKROLLNNVLPENGHVLDGKIGTTLGT--FLSSTALIRTS 463
Db 381 SFIQANVKDKTLLVPPTASVTRTLOGGLFPTWGD--TQKLTIDLQRTWMLSDVDFLKM 438
QY 464 ARAGYFFT-PENKKLGTFTIRQOAGYTVARDNADVPISGLMFRSGASSYRGYELDSIGLA 522
Db 439 GSSLVWRTYVQLQHRI---VTRLELGLWLTKNIERIPPALRPFPAGGDRSIRGYGKXIAPK 495
QY 523 GPNISVLPERALLVGSLEYQLPFTRTLSCAVFIDMGDAAANFKRMKLGKHSGLGVWRFSP 582
Db 496 NNAGKLVGGSRLTGSFQYQYVQVYVQVYVQVYVQVYVQVYVQVYVQVYVQVYVQVYV 555
QY 583 LAPFSDFDIAYGHSD---KKIRWHISLG 606
Db 556 VCAINFDIATPVDRDKNISQIYICLG 583

```

RESULT 10

```

Q8XZ13 PRELIMINARY; PRT; 765 AA.
AC Q8XZ13;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative outer membrane signal peptide protein.
GN RSC1412 OR RS05280.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangelot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646064; CAD15114.1; -.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000184; Bac_surfaq_d15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
DR PROSITE; P500061; ADH_SHORT; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 765 AA; 85878 MW; 378641FB9F247C68 CRC64;

```

Query Match

Best Local Similarity 10.5%; Score 332.5; DB 16; Length 765;

Matches 140; Conservative 89; Mismatches 245; Indels 97; Gaps 23;

Query Match	9.7%	Score	306.5	DB	16	Length	617
Best Local Similarity	23.3%	Pred. No.	3.3e-11				
Matches	140	Conservative	89	Mismatches	232	Indels	139
Gaps							25;
Qy	82	VLDKEOTG----	LAEEAPDNVKTMLRSKGVFSSKVS	---	TEKD	G-ATYVHI	PGPRTK
Db	79	VVGEQGEFRLEYLLSQEFQTRQALEPFGYVAPTIRIDAPQNDHITVVIYVDKGEVPR	138				
Qy	134	IANYGVAILGDILSDGNLAERYNALENMWOQPVGSDFDQDSWNSKTSVLGAVTRKGYPL	193				
Db	139	VRQAHVAMTGAQAQ----	HYLRDLEDFEKLGEIFNHPTYEASKVIRTRRLAERGYYD	194			
Qy	194	AKLGNTAAVNPDTATVDLNVVDSGR----	PTAFG--DPEITGTQRYPEQIVSCLARFQ	247			
Db	195	ADFTHRRVEVTRAAHADIDILWSGRYDMPVRFHYYD-----	FHEGLFNPLVYMD	247			
Qy	248	PGTPTDLDLLDFOALPQNGHYSGASQAQDFDLQGD--RVPVKVSVTEVRRHKLETCIR	306				
Db	248	EGSYFHEGKDLRLRESLTKLDYFSSIDIQKPEADPGSNVPDVKLERAKSKIYTAGIS	307				
Qy	307	LDSEYGLGCKTAYDYNYLNFNGYTGSVVYMDKMTYTTLLAAGISQP--RWYRGNYWTSNNVSY	365				
Db	308	YGESGAG-----	LRAGVERRYMNARGHKMNARLDY	338			
Qy	366	NRSTTQNLKRAFSGGI-----	WVYVRDAGIDARLGAEFFLAGRKIPGSDIDLG--	414			
Db	339	----AQLNLSLITAYQIPAFKWLGDWYI-----	FFARAYDEQTKYIDLNRVK	381			
Qy	415	-----NSHAT-----	MLTASWKROL-----	LNNVLHPE--NGHYLDCKTGT	448		
Db	382	LSAARSGOINRHLTATASLNALRCRWRYRADDGNTNVAQOSTLVIYPOLEASYVDVDDAT	441				

Query Match

Db	439	DNFGQVY-----GQLRWFLG-----AG--	450
Qy	470	FTPENKKIGTPIIRGOAGYTIVARDADVPGLMFRSGASSVRGYELDSIG--LAGPNGS	527
Db	457	--DNSRL--ILRGEGTTWTSDLVAMPPLSLRFFAGGSNSIRGYAFREVGPRTPKPDFX	510
Qy	528	VLPREALLVLSLEYQL-----PPTRRTLSGAVFHDMDAAANFRMKLHGSGGLGVRFSP	582
Db	511	ALGAKNVYTAASYEHYHLGGPW---GGAFVFDGSAFDD--RPDWHHTGIGFLWRSP	564
Qy	583	LAPSEFDIAYG--HSDKKIRWHISLG	606
Db	555	VGPRVDIAHGLNDPDQAQFLYIDIG	590
RESULT 15			
Q8YEI3	ID	PRELIMINARY; PRT; 623 AA.	
AC	Q8YEI3;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Outer membrane protein.		
GN	BMEI1895.		
OS	Brucella melitensis.		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Brucellaceae; Brucella.		
OX	NCBI_TaxID=29459;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=16M / ATCC 23456 / BIOTYPE 1;		
RX	MEDLINE=20020109; PubMed=11756688;		
RA	DelVecchio V.G., Kaparat V., Redkar R.J., Patra G., Mujer C., Los T.		
RA	Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,		
RA	Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.		
RA	Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,		
RA	Hasselkorn R., Kyrpides N., Overbeek R.;		
RT	"The genome sequence of the facultative intracellular pathogen		
RT	Brucella melitensis.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).		
KW	EMBL: AE009623; AAL53076.1; ..		
DR	Complete proteome.		
SQ	SEQUENCE 623 AA; 67227 MW; 07C43D0F6AD8F8FE CRC64;		
Query Match	9.2%; Score 290.5; DB 16; Length 623;		
Best Local Similarity	20.4%; Pred. No. 3.5e-10;		
Matches 140; Conservative 106; Mismatches 248; Indels 193; Gaps			
Qy	38	KNKSPDTSVKLRPK-FPVRI-----DTDSEIKDMVEHRLPLITQQOEVLDKEQ	87
Db	15	KQDPDI----IDPKTSVDVTTTGDRKNADGREADLKSVIEGASGLVSADAKPA--SGS	68
Qy	88	TGFLEAEAPD--NVKTMLRSKGYSSKVSL-----TE--KCQAYTVHTTPGP	130
Db	69	AGLLAKARGDYRRILSALYGEGRYGGTISIKVDGREANDIPPDTEIPNNAKVAITVDPGP	128
Qy	131	-----RKIANV-----GVALGDLSLDGNLAEEYRNALENKQCPVGS	168
Db	129	QFLSRSTAISNIAPPNGNRDRKVQTPEAGFA--PGQSAKSGTILKAERLAVEAWRO----	193
Qy	169	DFDQDSWENSNTSVLGAATRGYPYLAKLGHNTAAVNPDPTATVDLNVVVDSGRPIAFGDGE	228
Db	184	-----EGYAKRVCTGEDVADHADNRVSADIADLPGRKAYICPVCS	223
Qy	229	ITGTQRYPEQIVSGLARFPQGTPTYDLLDLLDFQALBQNGHYS--GASVOADFRLQGD-	285
Db	224	VVGVARMDPQVAMHTGLKPCQEYDPO---DIENAKRLGRMEVFRAMTPEADKIEPDG	280
Qy	286	RVPKVSYTEVRKHLETGIRLDSYGLGGKIADYYNLFNKGY-----IGSV--	333
Db	281	SLPITLNVQERKPRRFGEABYSYIDGFGVTSYMMHRNLLGRGERURFDAKVSIGGSQD	340
Qy	334	-VWMDKYETTLAAGISOPNRYRGN-----YNTSNVSYNRSTTO-----	371

```
Db 341 NSFDPKNTYLLGASFAKPGVYTPDQFVATLDKAREVLDAYTETSIINAKTGFQIFSD 400
Qy 372 -----NLEKRAFSGGTYWRD--RAGIDARLGAELAEGRKIPGSDIDLGNSHATMLT 422
Db 401 LSGALYANASOGHFVDDVFGKRDRTTAGLEGNL-----LYDSR----- 438
Qy 423 ASMRQLLNVLHPENGHYLDGKICT-----TLGTFSLSTALIRTSARAGYFFTPENKKLG 478
Db 439 -----NKKPDSSGFYLVGNIOFPVEFYHGNFATR-----FTAEGRTYH 477
Qy 479 TF-----IIRG--QAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGNGSVLP 530
Db 478 GFGQTDRLVLAGRLKVGSIIVGSIADLPSPQLFLAGGGSVRGYCYRIGVSAAGNIGI 537
Qy 531 ERALLVGSLEYQLPPTRTLSCGAVFDMG----DAANFKRMKLGKSGGLGVWRFSPPLPF 586
Db 538 GRSVLEANGEVKTRITDISIGAVFVADAGYVGEKSPDPFSE--OMRVGVGGGLYLTSLGPI 596
Qy 587 SFDIAY-----GHSDDKKIRWHISLQTRF 609
Db 597 RLDVAVPLNRRSGDPNYGFFYVIGQAF 623

RESULT 16
Q8YUR6 PRELIMINARY; PRT; 833 AA.
AC Q8YUR6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical protein Alr2269.
GN ALR2269.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73968.1;
DR InterPro; IPR000184; Bac_surfaG_d15.
DR Pfam; PF01103; Bac_surfaceAg; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 833 AA; 89633 MW; 7F52B1E52FEA1BE9 CRC64;

Query Match 8.8%; Score 279; DB 16; Length 833;
Best Local Similarity 22.9%; Pred. No. 3e-09;
Matches 153; Conservative 82; Mismatches 234; Indels 200; Gaps 26;

Qy 81 EVLQKEQTGFLEAPDNVKTMLRSK-----GYESSKVSITEKDG 120
Db 225 EVLVPSGGQTLPELETQVYVIRTPGRTTTRSQLQEDINAIFGTGFFSN--VOASPEDT 283
Qy 121 AYTVHIT-----PGP----RTKI--ANVG-----ATLGDILSDGNLAEYR 156
Db 284 PLGVRVSPVQPNVPSLRSKVEIQANPGTNVPSVLPQATADEIFRAQYKILNRLDQEGIK 343
Qy 157 NALENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAK--IGNTRAANVPOTATVDLVN- 214
Db 344 ELTRYQD-----QGVYLANVVGAPQVSEN---GVVTLQVA 376
Qy 215 -----VVDSGRIAPGDFEITGTQRYPEQIVSGLARFOPGTPYDLDL 256
Db 377 EGVVENISVRNKGQDVNECGQPI-----RGRTQDYIITREVELAPQGVFNRT 427
```

```
Qy 257 LLDFQQALEQNGHYSGASVQADFORLQGRVPPVKVSVTEVRRHKHLETGIRLDSYGLGCK 316
Db 438 VOKDLQRVFGTGLFEDVNVSLD--PCTDPTKVNVVVVVVERSSGSTAACAGISSSSGLFCT 486
Qy 317 IAYDYNL-----FNKGYIGSVVWMDKYEYTTLAAGISQPRN 353
Db 487 VSYQOONLNGRNQKLGAEVOLGERELLDLRFTRDPWIGG-----DPYRTSYTANIFRRRS 541
Qy 354 YR-----GNYWTSN-----VSYNRSTTQNLKRA---FSGGI 382
Db 542 ILSLFDGKDEDIRFDGPNPNDTNGQDRPVTRTLCGGVYFTFRLPSANPERAEWTASAGL 601
Qy 383 WY-----VRDRAGIDARLGAELAEGRK----IPGSDIDLGNSHATMLTASMKROLLNVVL 434
Db 602 QYQVSTRDADGNLRKDCGAVFDDNGNRTSEIVPLSFSGTGEDDLLLVQGAQDLRNPL 661
Qy 435 HPENCHYLDGKIGTGTTLGTFSLSTALIRTSARAGYF-----FTPENKLGKTFPIRGQAGY 488
Db 662 QPTSGSFL--RFGVDQSVPGVSGNIFLTRFRGYSQYLPVAKFTGFSKGPETIAFNIOGQT 719
Qy 489 TVARDNADVPGLMFRSGGASSVRGYELDSIGLAGNGSVLPERALLVGSLEYQLPPTRT 548
Db 720 VL-----GDLPPYEFTLGGNSVRGYEGALGSG-----RSFQASVEYRFPVFSV 766
Qy 549 LSGAVFHDMDG-----AAANFKRMKLGKSGGLGVWRFSPPLAFPSFDIAYG--HSDKKIR 600
Db 767 VSGALFFDVGSDLGCTSTRTAENVLNKSGSGYGVGLGVQSPGLGPIRID--YGINDGDSDR 824
Qy 601 WHISLQTRF 609
Db 825 INFGIGERF 833

RESULT 17
Q9AYW6 PRELIMINARY; PRT; 628 AA.
AC Q9AYW6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein CCI603.
GN CCI603.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CBI5;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Deboy R.T., Newton W.C., Stephens C., Phadke N.D., Ely B.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Haft D.H.,
RA Otterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005835; AAK23582.1; -.
DR TIGR; CCI603; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 628 AA; 68203 MW; 92DAD1722CF82295 CRC64;

Query Match 8.6%; Score 272.5; DB 16; Length 628;
Best Local Similarity 23.2%; Pred. No. 4.9e-09;
Matches 138; Conservative 92; Mismatches 285; Indels 81; Gaps 19;

Qy 60 QDSEIKDMVEEHLPLITQQEEVLDKQETGFLAEAPDNVKTMLRSKGYFSSK--SLTE 117
Db 68 EDRLRDAIQRAL---SDSKQPPRSRRARRARAGEDATAVLRAEGYAYTVPEPDYTE 124
Qy 118 KGAYT-VHITPGPRTKTANVGVAITLSDGNLAEYRVALENWQOPVGSDFDQDSWE 176
```

```
Db 125 GPPRAIVRITPGFAFLADPHDWSGPPDEGVR-----QRAVAAMLTEGEP 173
Qy 177 NSKTSVLCAVTRKGYPLAKLGNTRAAVNPDTATVD-----LNWVDSGRPIAFGDEI 229
Db 174 GRSADVGAEGRIVAQAKLGYADVAEPREVVDHADRTVPRFIRIMAGELVRLNGVDV 233
Qy 230 TGTQRYPEQIVSGIARFQPGPYDLDDLLDQQAQLEONGHYSYGASVQ---ADFRLQGD 286
Db 234 VTKGRTNPEWGRILAPWAGVDYDPEDVAELERRLDRTAVYDSISVSLAGTDKASAEGR 293
Qy 287 VPKVSVTEVRHKLETCIRLDSEYGLGCKIAYDYNLFNKGYIGSVVWMDKYEITLLAA 346
Db 294 -PVVYTLSDRRARTIELGAGYSTSEGAGVDARWIRYNRQKRADTTTYALRFPAKLEORLGA 352
Qy 347 GISOPRNYRGNYWS-----NVSYNRSTTONLEKRAFSGGIWYVRDRAGIDAR-LGAE 398
Db 353 EISLP-----HWRRPQOTKLNSVFRNDTDAVNETGATVGVDLTRRRQTTATRTFGVS 406
Qy 399 F-----LAEGRKIPGSDIDLGNSHATML-TASWKROLNNVLPENHGYLDG 444
Db 407 FDLSTQKEQVNRNGLIAGRK-----NLATLAGLAAYAVWDFSDDILDPKRGWRLET 457
Qy 445 KIGTTLGTFSLSTALIRTSARAGYFFTPENKKGTFI-IRQAGYTVARDNADVPGLMF 503
Db 458 RAEPYVAGDTSVPLYKLAGO-GSAYLPFGKODSTVLAARYKLGAIGLAGLDDVPASRRF 516
Qy 504 RSGGASSVRGYELDSIGLA-----GPNGSVLPERALLVGSLEVQLPFTTLTSGAVFHDMD 559
Db 517 FSGGGSVRGYAYQAIQPRLSNDTPQGGI-----SLVETSFVRQKITDRWSGVAFVDAGA 572
Qy 560 AAANF--KRMKLKIGSGLVGRWFSP LAPPSFDIA-----YHSDKKIRWHISLGTRF 609
Db 573 IGTHTPOREDFRAGLGVRYDLGFGPIRADIAAPLGRKGDPKFQIYLSIGOSF 628
```

RESULT 18

```
ID P95359 PRELIMINARY; PRT; 792 AA.
AC P95359;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein.
GN OMP85.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RA Manning D.S., Reschke D.K., Judd R.C.;
RT "Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis are similar to Haemophilus influenzae D-15-Ag and Pasteurella multocida Omp87."
RL Microb. Pathog. 23:0-0(1998).
DR EMBL: U81959; AAC17600.1; -
DR InterPro: IPR000184; Bac_surfaAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1
SQ SEQUENCE 792 AA; 87868 MW; 90E32D24AA0513D8 CRC64;
```

Query Match 8.2%; Score 260.5; DB 2; Length 792;
Best Local Similarity 22.6%; Pred. No. 4.1e-08;
Matches 139; Conservative 86; Mismatches 242; Indels 149; Gaps 28;

```
Qy 97 DNVTMLRSKGYF-----SSKVSLETKDCATVTHIT--PGPRTKIANGVAILGDSG 149
Db 223 EKVTFDYQNVGYDFRILDTDIOTNEDKTRQTIKTIHGGFRWKGYSI-----EG 274
Qy 150 NLAEYRNAL------NNQOPVGSDFDQSDNSWNSKTSVLGAVTRK-----GYPLAKLG 197
Db 275 DTNEVPKAELEKLLTMKPGKWE-----RQQMTAVLGEIQNRMGSGAYVSEI- 322
```

```
Qy 198 NTRAAVNPDTATVDLNWVVDGRPIAFGDFEITGQRYPEQIVTSGIARFQPGTTPYDLDL 257
Db 323 SVQPLNAGTKTYDVFVLIHIEPGRKIYVNEIHITGNKTRDEVVRRELQMESAPYDTSKL 382
Qy 258 LDFQALQONHYSYGASVQADFDRLOG--DRVPKVSVTEVRHKLETCIRLDSEYGLG 315
Db 383 QRSKERVLLGYFD--NVQFDVPLAGTPDKVDLNMSLTERSTGSLDLSAGWVODTGLVM 440
Qy 316 KIAVDYVNLNFKGYIGSVVWMDKYEITLLAAGISQPRNYRGNYWS-----NVSYN----- 366
Db 441 SAGVSQDNLFGTG--KSAALRASRSKTTLNGSL-----FTDPIYFTADGVSLGDIYIGRAF 494
Qy 367 -----RSTTONLEKRAFSGG-----IWYVRDRAGIDA-----RLGAFLAE 402
Db 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNKAPKRYADFIKQ 554
Qy 403 GRKIPGSDIDLGNSHATML---TASWKROLNNVLPENHGYLDGKIGTTLGTFSLSTAL 459
Db 555 YGKTDGAD-----GSFKGLLYKGTVGWGRNKTDLSALWPTRG-YLTG-----VNAPI 599
Qy 460 IRTSARAGYFFTPENK-----KLGTFIIRQAGYTVARDNA-DVPSGLMFRSGGASV 511
Db 600 ALPGSKLOYYSATHNOTWFFPLSKTFTMLGGEVGIAGGYGRTKEIPFENYGGGLGSV 659
Qy 512 RGYELDSTIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FRTLSGAVFHDMD 559
Db 660 RGYE---SGTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLPADAGS 716
Qy 560 -----AAANFKRM-----KLKHSGLGVGRWFSP LAPPSFDIAY-----G 593
Db 717 VMDGRTYTAAGENGNNKSVYSENAHKSFTTNELRYSGAGAVTWLSPLGPMKFIYAYPLKK 776
Qy 594 HSDKKIRWHISLGTRF 609
Db 777 PEDEIQRFQFQLGTF 792
```

RESULT 19

```
Q9K1H0 PRELIMINARY; PRT; 797 AA.
ID Q9K1H0
AC Q9K1H0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein Omp85.
GN NME0182.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AE002375; AAF40639.1; -
DR TIGR: NMB0182; -
DR InterPro: IPR000184; Bac_surfaAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Complete proteome.
SQ SEQUENCE 797 AA; 88436 MW; B68BFC5A43D22EE8 CRC64;
```

Query Match 8.0%; Score 253; DB 16; Length 797;
Best Local Similarity 21.9%; Pred. No. 1.3e-07;

Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

Qy 97 DNVTMLRSKGYF-----SSKVSLTEKDAGYTVHIT--PGPRTKIANGVAILGDILSDG 149
: : | | | : : | : : | : : | : : | : : | : : |
Db 223 EKVDFYQNNCYFPDIRLTDIQTNEDKTQTIKITVHEGGFRGWKYSI-----EG 274
: : | | : : | : : | : : | : : | : : | : : | : : |
Qy 150 NLAETYYRNALB-----NWQPVGSGFDODSWSNKSATVLGVATR- ---GYPLAKLG 197
: : | | : : | : : | : : | : : | : : | : : | : : |
Db 275 DTNEVPKAELEKLTMKPKNWE-----RQMATAVLGETQRMGSGAYSEI- 322
: : | | : : | : : | : : | : : | : : | : : | : : |
Qy 198 NTAAVNPDATVDLVNVVVDSGRPIAFGDPFEITGTQRYPEOIVSGLARFOFGTTPVDLLD 257
: : | | | | | | | | | | | | | | | | | | | | | |
Db 323 SVQPLPAEIAETKTYDVFLHIEPRKIYNEIHITGNKRTRDEVRRELROMESAPDTSKL 382
: : | | : : | | : : | | : : | | : : | | : : | | : : |
Qy 258 LDFOALEFNQHYSYGASQAQDFDLQG--DRPVVKVSVEYRKHKLETIGRLDSEYGLGG 315
: : | | : : | | : : | | : : | | : : | | : : | | : : |
Db 383 QRSKEVELLYGF--NQVFDAVLAGTPDKVDLNMSLTRSTGSLDLSAGWVQDTGLVM 440
| | | | : : | | : : | | : : | | : : | | : : |
Qy 316 KIAVDYNLFNKGIYGSVWMMDKYETTL-----AAGISOPRNYRGNNW---- 359
| | | | : : | | : : | | : : | | : : | | : : |
Db 441 SAGVSQONLFTGT--KSAALARSRSKTTNLNGSLTSFTDPYPTADGVSLCYDYVGKAFFPRK 498
: : | | : : | | : : | | : : | | : : | | : : |
Qy 360 -TSNVSYNRSTONLEXRAPSGGIWYPDRAGIDA-----RLGAEFFLAEGRKI 406
: : : : | | | | : : | | : : | | : : | | : : |
Db 499 ASTSIKQYTKTTAGAGIRMSVPVTEYDRVNFGLVAEHJTVNTYNKAPHYADF TKKYGT 558
: : | | : : | | : : | | : : | | : : | | : : |
Qy 407 PGSDIDLCSHATML--TASHKROLLNNVLPENGHYLDGKIGTGTFGLFSSTA LIR TSA 464
| | | | : : | | : : | | : : | | : : | | : : |
Db 559 DCTD---GSFKCWLXKGTVCHGRNKTSALWPTRG-YLTG-----VNAELALPGS 604
: : | | : : | | : : | | : : | | : : | | : : |
Qy 465 RAGYFFTPEPK-----KLCTFIIRGOAGYTVARDNA-DVPSGLMFRSGASSYRGVEL 516
: : | | : : | | : : | | : : | | : : | | : : |
Db 605 KLOQYSATHNQTFEPFLSKTFTLMGGEBVIAGCVGRTKEIPFENFYGGGLSGSVRGYE- 663
: : | | : : | | : : | | : : | | : : | | : : |
Qy 517 DSIGLAGPN-----GSVLP-----BRALLVGSLEQLP---FTRLISGAVFHDMGD----- 559
| | | | : : | | : : | | : : | | : : | | : : |
Db 664 --SCTLGPVKYDEYGEKISYGNKKANVASALLEPFMPGAKDARIURLSLFADAGSVNDGK 721
: : | | : : | | : : | | : : | | : : | | : : |
Qy 560 -----AAANFKRM-----KLKHGSLGVWFSP LAPSFEDIA Y ----G 593
: : | | : : | | : : | | : : | | : : | | : : |
Db 722 TYDDNSSSATCGRVQNIYGACHTHSKTFTELRLSYAGGAVTWLSP LGPMKYSAY PLKKK 781
: : | | : : | | : : | | : : | | : : | | : : |
Qy 594 HSDKKIRWHISLGRTRF 609
: : | | : : | | : : | | : : | | : : | | : : |
Db 782 PEDETQRFFQLGTTTF 797

RESULT 20

Q9JX31 PRELIMINARY; PRt: 797 AA.

ID AC Q9JX31;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
Outer membrane protein OMP85.
GN OMP85 OR NMA0085.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_taxid=65699;
[1]
SEQUENCE FROM N.A.
SYRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
MEDLINE=20222556; Pubmed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR ENBL; AL162752; CAB83401.1; -.

```
RL Microb. Pathog. 23:0-0(1998).
DR EMBL; AF021245; AAC17599.1; -.
DR InterPro; IPR000184; Bac_surFag_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
SQ SEQUENCE 797 AA; 88539 MW; CF911B5F70B9999CF CRC64;

Query Match 7.9%; Score 251; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 1.7e-07;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

Qy 97 DNVTMLRSKGYF-----SSKVSLETKDGAFTVHIT--PGPRTKIANVGVAIGDILSDG 149
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
223 EKVYDFQNNNGYDFRILDTDIQTNEDTKTKITIKITVHEGGRFMGKYSI-----EG 274
Qy 150 NLAETRYNALD-----NMQQPVGSDFDODSWNSKTSVLGAVTRK----GYPLAKLG 197
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 DTNEVPRAELEKLTMPKWE-----RQQTAVLGEIQNRMGSAGYAYSEI- 322
Qy 198 NTRAAVNDPTATVDLNVVDSGRPIAGDFEITQRYPEQIVSGLARFQPTPYDLDLL 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 SVQPLPNAETKTVDVFLHIEPGRKIYVNEIHTGNKTRDEWRRELQMESAPYDTSKL 382
Qy 258 LDFQALQONGHYSGASVQADFRLQ--DRVPVKVSVTEVKRHKLETGIRLDSEYGLGG 315
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 QRSKERVLLGYFD--NVQFQDAVPLAGTPDKVDLNMSLTERSTGSLDSAGWQDTGLVM 440
Qy 316 KIAYDYNLNFKNKYGIVVMDMKYETTL-----AAGISOPRNYRGNTW----- 359
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 SAGVSQDNLFTG--KSAALRASRSKTYTLNGSLSFDPYFTADGVSLGYDVYKAFDPK 498
Qy 360 -TSNVSYNRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAFLAERKI 406
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
499 ASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYKAPKHVADFIKKYKT 558
Qy 407 PGSDIDLGNSHATML--TASHKROLLNNVLPENGHYLDGKIGTTLGTLFSLALRTSA 464
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
559 DGTD--GSGFKWLYKGTGVGRNKTDSALWPTRG-YLTG-----VNAEIALPGS 604
Qy 465 RAGYFFTPENK-----KLGTFTIRGQAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
605 KLOYSATHNQTWFFPLSKTFTMLGGEVGIAGGYGRTEKIPFPENFYGGGLSGVRGYE- 663
Qy 517 DSIGLAGPN-----GSVLP-----RALLVGSLEYQLP---FTRLSCGAVFHDMDG- 559
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
664 --SGLTGPVKVYDEYGEKISYGNKKANVSAAELFPMPKAKDARTVRLSLFADAGSVWDGK 721
Qy 560 -----AAANFRM-----KLKHGSLGCVWFSPFLAPFSDIAY-----G 593
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 TYDNSSSATGGRVONIYGAGNTHKSTFTNELRVSGAGVATWLSPLGPMKFRAYPLKKK 781
Qy 594 HSDKKIRWHISLGRF 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
782 PEDEIQRFQFOLGTF 797

RESULT 22
O67326 PRELIMINARY; PRT; 778 AA.
AC O67326;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein.
GN OMP OR AQ_1300.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,
[1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT *Genome sequence of Yersinia pestis, the causative agent of plague.;
RL Nature 413:523-527(2001).
DR EMBL: AJ414146; CAC89894.1; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 795 AA; 87838 MW; 8C4581328D38F821 CRC64;

Query Match
Best Local Similarity 21.2%; Score 246; DB 16; Length 795;
Matches 142; Conservative 105; Mismatches 273; Indels 150; Gaps 30;

QY 47 VLKPKFPVRIDTQSEIKDMVEEHLPLITQOEEVLDEQGTGLAEAP----- 96
DB 163 VDLKLVTEGVSAKIQOI-NIVGNH-----SFTTDELISRQ-----LRDEYPMWNVGDRKY 214
QY 97 -----DNVKTMLRSKGYF-----SSKVSILT-EKDGAY-TVHTPGPRTKIANVGVA 140
DB 215 OKQKLAGDLETLSRFLDRGYARENFIDSTQVSLTPDKKGIYVINTITEGPQPK----- 267
QY 141 ILGDILSDGNLAIEYYRNALENNQQPVGSDFDQDSWNSKTSVLGAVTRKGYPLAKIGNTR 200
DB 268 -LNSVIVSGNLACHGSEAEKTKIEPGELFNGSKVTRMEDDIKKMLGRGYAYPRV-VTQ 325
QY 201 AAVNPDTATVDLNVVDSGRPTAFCDFEITGTQRYEQIVSGLARQPQPTPDLDLLDF 260
DB 326 PEINDDKTKVHLINVDAGNRYVRHIREGNDTSKDSVLREMRGMEGAWLGNDOVEAG 385
QY 261 QOALEONGHYSAGSQAQDFDRLOG--DRVPVKVSVTEVKRHKLETGIRLDSYGLGKIA 318
DB 386 KERLNLRYGFE--TVDVETQRPVGAADLDVTVYKVKERTNTGSLNFGIGYTESGSFQVG 443
QY 319 YDYINLFNKGYSVGVVWMDKVTETTLAAGISQPRNRYGNYWT-----SNVSYNRSTQ 371
DB 444 VOODNMLGTGNTVINGTNGYQTYAEFTLMDP-----YFTVGVSLGGRIFYNDFRAD 497
QY 372 NLEKRAFSGGIYVVRDRAGIDARLGAELAEGRKIPGSDIDLGNH--ATML--TASMK 426
DB 498 NAD---LSG---YTNSSYGADGTGLGP--INENNSL---RVGVYVHNDLMDLPQVAMWR 547
QY 427 RQLNNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYF----- 469
DB 548 --YLESV--GERPGY--DGREGTDDFTLNLGWTYNNLORGFPPTSGVKSSVNTKITVPG 602
QY 470 -----FTPENKLGTFII-RQAGYTVARNADNVPGLMFRSGGASSVRGY 514
DB 603 SONEFYKVTFDTSAYOPLNEDRSWLLGRGLGYDGIGSKEMPEFENFYAGGSVTRGF 662
QY 515 ELDSIGLAG--PNG-----SVLPERALLVGSLEYQLP-----FTRLSGAVFHD 556
DB 663 RNNIGPKAAYTANGATVTNSTDAVGGNMAVASLELITPTPFISEKYSNSVRTSIFID 722
QY 557 MGDA-----AANF-----KRMKLGHSGLGVRMFSPLAFSFDIAYG-----HSD 596
DB 723 SCTVMDTNMENTAKTRAAGIPDYGKASNIRVSAGVALQWMSPLGPLVFSYAKPVKDYEGD 782
QY 597 KIRWHISLG 606
DB 783 KSEQFOFNIG 792

RESULT 24
Q98K30
```

```
ID Q98K30 PRELIMINARY: PRT; 617 AA.
AC Q98K30;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein mill1662.
GN MLL1662.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFE303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti".
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002998; BAB48984.1; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR InterPro: IPR003975; Shal_channel.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR PRINTS: PR01497; SHALCHANNEL.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 617 AA; 65759 MW; 83FB8C176201A444 CRC64;

Query Match
Best Local Similarity 20.9%; Score 240.5; DB 16; Length 617;
Matches 133; Conservative 109; Mismatches 284; Indels 111; Gaps 24;

QY 21 APAADLSEKAKAGFALFKNK-----SPDTSVKLKP-KFPVRIDTQSEIKDMVE--EHL 73
DB 4 ATALVAESPPAAFAELFGIKLWSSNDEADIVDLRYAVTITAPDAD-KDLVKRLNAS 62
QY 74 LITQOQEEVLDEQGTGLAEAPDNVKTW--LRSKYFSSKYSLTEKDGAYTVHTP--- 128
DB 63 ALKGDEDRPV-SGSLGLMAKARSDRELVAALFADARYEGVVTIT-IDKPLDELPPDAE 120
QY 129 --GPRTKIANVGVAI-----LGDILSDGNLAIEYYRNALENNQQPVGSDFDQDSWNSKTS 181
DB 121 FXGPOPIPTVINIAGPKFTLGTIRLEGDAAGLMS---ADYGLISGDAGSAGVAKAEAL 177
QY 182 VLCAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPE---Q 238
DB 178 IVRTLKEQGRPLAKVTRDRIVADHATSTLDVTLTVAAGPVAGYGATTVEGTEKVDRTFE 237
QY 239 IVSGLARFOPGTPYDLDLDDLLDQALEQNGHYSAGSQAQDFDRLOG-RVPVKVSVTEVK 297
DB 238 YMTGLARGQYSPQELSDARDRLALEV---FNSVTFK-EADKLDADGNIPIGVQVSEK 293
QY 298 RHKLETGIRLDSYGLGKIAYDYNNLFNKGY-----IGS----- 332
DB 294 PRYFGLGGTFSNTEGILEGSGYGNHRLFGHAEKLRIDGALISGNSNLSDLNYNAGIMFE 353
QY 333 ---VWMDKYEYTLAAGISQPRNRYGNTWTSNVSYNRSTTQNLKRAFSGGIWIYVDRRA 389
DB 354 KFCVIGPASKFTFAGKTVLEHPDAY-----DHFSYKGST-----GLSYELDK- 395
QY 390 GIDARLGAFLAEGRKIPGSDIDLGNSHATMLTASWKROLLNNVLHPENGHYLDGKIGT 449
DB 396 ---QQTVAEVALDYSRI-----TDAFKHHYLIASVPLOQVYDN-----RDSRLNPT 440
QY 450 LG-----TFLSSTALIRTSARAGYFFTPENKILGTFIIRQAGYTVARNADVP 498
DB 441 RGRFRLAYAEPSYVDIMSGAFLKLGEGSAYQSLDTASKFVLAERVAIGSIVGTGLQNPV 500
QY 499 SGLMFRSGGASSVRGYELDSIGLAGPNSVLPERALIVGSLEYQLPFTTTLSCAVFHDMP 558
```



```
Db 504 --YTNQSYGICSGFLGPPINNENNSLNFLNLYIHNSLSDMLPQVAMWRYLRSMGEKPDLESK 561
QY 397 AFLAEGRRKIPGSDIDLGNHSHATMLTASWKQLLNVLHPENG--HYLDGKIGTGTGLTFL 454
Db 562 AEFKADD-----FALTMGWYNNLDRGFFPTSGVKSTLNGKV-TIPG--- 602
QY 455 STALIRTSARAGYFTTTPENKKGITFIIRGQAGYVARNADNVPGLMFRSGGASSVRGY 514
Db 603 SDNEFYKVTLDTSAYPIINDRTWILGRSLRGYDGLGGKELPFYENFYAGSGSTVRGF 662
QY 515 ELDSIG-----LAGPNSVLPERALLVGSLEYQLP-----FTRTSLGAVF 554
Db 663 RSNNGIPKAIYLYKDGSPKESPSRDVAGGNAMAYALELITPTFLDSKYSNSVRTSFF 722
QY 555 HDMG-----DAAA-----NFKRMKMKHSGSLGVRFSPPLAFSFDIA 591
Db 723 IDSGVTWTDNRNDSAVMKSKGIPDYSKGNIRVSAGIALOWNSPLGLPLVFSYA 775

RESULT 27
Q9KPW0 PRELIMINARY; PRT; 803 AA.
AC Q9KPW0:
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Surface antigen.
GN VC2252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Hodelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004297; AAF95396.1; -.
DR TIGR; VC2252; -.
DR InterPro; IPR000184; Bac_surfaAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match
Best Local Similarity 7.2%; Score 227; DB 16; Length 803;
Matches 135; Conservative 95; Mismatches 225; Indels 168; Gaps 31;

QY 84 DEQGTGFLAEEADPVKVTMLRSKGYFSSKSLT-----EKDGAY-TVHITPGPRTKIAN 136
Db 212 DRYKKQVLAGDI-EALRTYLDRLGYLKFQVOSTQVAISPDKKGVITLNLNAGEPYTVSK 270
QY 137 VCVAILGDILSDGNLAERYRNALENWQPVGSDFDQDQSWENSKTSVLGAVTRKGYPLAKL 196
Db 271 --VQFRGELM--GKEAEP--TSLIPFE--IGETYNGSAVTRLEESKVKVLGSGVAYPQV 322
QY 197 GNTRAAVNPDTATVDNLNVVDVSGRIAFGDFEITGQRYPEQIV----- 240
Db 323 -RTIPEFDEKQOQVSLVHVEAGKRVYVDIRFVGNSTRDVLREMRQMEGSLNSKD 381
QY 241 --SGLARFQ-----PCTPYDLDDLDDFQALEQN-----CHYGSASVQ 276
Db 382 IETGKTRNLRLGFFETVEVQTVRVPGSEDOVDLVYTVSKEANSNGNMGVFGYGTSGVSFQ 441
QY 277 ADFDR----LOGDRVPKVSVEVTRHKLETCGIRLDSEY-----GLGCKTAYDY 322
```

```
Db 442 VGLQDNFLGSGNRVGVNAMINDYQKN-----LTLEYRDPYWNLDGVSGLGKV---FY 491
QY 323 NLFNGKYGISVVWMDKYEETTLAAIGISOPRNYRGVWTSNVSYNSTTQNLKRAFSGGI 382
Db 492 NOFEASEAGIVDTYTESYGTSLTWGF--PFD-----ELNRFEGCI 529
QY 383 WYVRDRAG-IDARLGAEE-FLAEGRRKIPGSDIDLGNHSHAT---MLTASWKRLQNLNVLHPE 437
Db 530 GYTHNKNIGNLTPYLOVENFLA----AQASNIDSGGNLLTDDPFDINLSWTRNLNNSYFPT 585
QY 438 NGHYYLDGKIGTGTFLTSLTALIRTSARAGYFTTTPENKKGITFIIRGQAGY-----TVA 491
Db 586 AGNHQRAEYKMTVPG--SDAQYFKLQYDVQRQYFPLTKKHEFTLLRLGLGYNGYGQTDG 643
QY 492 RDNADVPCLMFRSGGASSVRGYELDSICLAGP-----NGS-----VLPERAL 534
Db 644 KDNL-FPFYENFYAGFTSLRGFSNS---AGPKAVRYDYSNGNSGSDTATDSDVGGNAI 699
QY 535 LVGSLEYQLP-----FTRTSLGAVFHDGMG---DAAANFKRMKMLKHGS----- 573
Db 700 ALASVELIVPTPFASEEARNQIRTSIFYDMSASWMDTEPDY-RGKADYGNQYVYDSDPTN 758
QY 574 -----GLGVRFSPPLAFSFDIA 591
Db 759 VRSSYGVALQWVSPMGPLVFSLA 781

RESULT 28
Q8Z9A3 PRELIMINARY; PRT; 803 AA.
AC Q8Z9A3:
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Outer membrane protein.
GN YAEF OR STY0247.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skellon J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAD08682.1; -.
DR InterPro; IPR000184; Bac_surfaAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89466 MW; BBF8670070E67628 CRC64;

Query Match
Best Local Similarity 7.1%; Score 226; DB 16; Length 803;
Matches 136; Conservative 102; Mismatches 284; Indels 154; Gaps 25;

QY 47 VKLKPKFPVRIDTQDSEIKDWVEEHLPLITQOOEEVLKDEQGTFLAERAP----- 96
Db 163 VDLKLVFQEGVSAKQOI-NIVGNH-----AFSTEELISHIFQ---LRDEVPMNMYVGDY 214
QY 97 -----DNVKTMLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGA 140
Db 215 QKQKLAGDLETURSYLDRGYARFNIDSTQVSLTPDKKGIYITVNIITEGDQYKLSGVQS 274
```


RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Bendlak L., Mamelak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smba gene, a suppressor of
RT the muka null mutant of Escherichia coli.";
RL J. Bacteriol. 174:7517-7526(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94240115; PubMed=8183897;
RA Janosi L., Shimizu I., Kaji A.;
RT "Ribosome recycling factor (ribosome releasing factor) is essential
RT for bacterial growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91317739; PubMed=1860827;
RA Shimizu I., Kaji A.;
RT "Identification of the promoter region of the ribosome-releasing
RT factor cistron (frr).";
RL J. Bacteriol. 173:5181-5187(1991).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90062117; PubMed=2684966;
RA Ichikawa S., Kaji A.;
RT "Molecular cloning and expression of ribosome releasing factor.";
RL J. Biol. Chem. 264:20054-20059(1989).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=86008268; PubMed=2995358;
RA Icho T., Sparrow C.P., Raetz C.R.H.;
RT "Molecular cloning and sequencing of the gene for CDP-diglyceride
RT synthetase of Escherichia coli.";
RL J. Biol. Chem. 260:12078-12083(1985).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91123198; PubMed=1991717;
RA Hirvas L., Koski P., Vaara M.;
RT "The ompH gene of yersinia enterocolitica: cloning, sequencing,
RT expression, and comparison with known enterobacterial ompH
RT sequences.";
RL J. Bacteriol. 173:1223-1229(1991).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91100302; PubMed=1987124;
RA Dicker I., Seetharam S.;
RT "Cloning and nucleotide sequence of the fira gene and the fira200(ts)
RT allele from Escherichia coli.";
RL J. Bacteriol. 173:334-344(1991).

RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=88058790; PubMed=2824445;
RA Crowell D., Reznikoff W., Raetz C.;
RT "Nucleotide sequence of the Escherichia coli gene for lipid A
RT disaccharide synthase.";
RL J. Bacteriol. 169:5727-5734(1987).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88058791; PubMed=3316192;
RA Tomaszewicz H.G., Mchenry C.S.;
RT "Sequence analysis of the Escherichia coli dnaE gene.";
RL J. Bacteriol. 169:5735-5744(1987).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88139188; PubMed=3277952;
RA Coleman J., Raetz C.;
RT "First committed step of lipid A biosynthesis in Escherichia coli:
RT sequence of the lpxA gene.";
RL J. Bacteriol. 170:1268-1274(1988).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92380982; PubMed=1355089;
RA Li S., Cronan J.;
RT "The genes encoding the two carboxyltransferase subunits of
RT Escherichia coli acetyl-CoA carboxylase.";
RL J. Biol. Chem. 267:16841-16847(1992).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094229; PubMed=1688424;
RA Zhou Z., Syvanen M.;
RT "Identification and sequence of the drpA gene from Escherichia coli.";
RL J. Bacteriol. 172:281-286(1990).
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90370122; PubMed=2203971;
RA Eriani G., Delarue M., Poch O., Gangloff J., Moras D.;
RT "Partition of tRNA synthetases into two classes based on mutually
RT exclusive sets of sequence motifs.";
RL Nature 347:203-206(1990).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93094132; PubMed=1459951;
RA Gervais F.G., Drapeau G.;
RT "Identification, cloning, and characterization of rcsP, a new
RT regulator gene for exopolysaccharide synthesis that suppresses the
RT division mutation fts284 in Escherichia coli K-12.";
RL J. Bacteriol. 174:8016-8022(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94124004; PubMed=7904973;
RA Allikmets R., Gerrard B., Court D., Dean M.;
RT "Cloning and organization of the abc and mdl genes of Escherichia
RT coli: relationship to eukaryotic multidrug resistance.";
RL Gene 136:231-236(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;


```

Db 596 DLS---FYKIDYR-GQVFAPLTDNY-TMRHTELGVDGYGSTERLPFFYENYAGGNSV 650
QY 512 RGYELDSIG-----LAGPNSVL-----PE-----RALLVGSLE--Y 541
Db 651 RGFKSTLGRSTPSVARPNPDGTPMKNOGSDSKGRYTDQDQPEAPGNNILITGGALLF 710
QY 542 QLPFT---RTLSGAVFHDMDA-----AANFKRMKLKHGSGIGVRFWSPLAPP 586
Db 711 PLPFPVKDQRLRTVLPMVDVSGTFDTCPTKTTNCIDGIKTDNLASSVGVGLTWTALGPL 770
QY 587 SFDA 591
Db 771 SFSLA 775

RESULT 32
Q8YMP0 PRELIMINARY: PRT: 676 AA.
AC Q8YMP0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Alr4893.
GN ALR4893.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL Ana Res. 8:205-213(2001).
DR EMBL; AP003597; BAB76592.1; -.
DR InterPro: IPR000977; DNA_ligase.
DR PROSITE: PS00697; DNA_LIGASE_A1; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 676 AA; 72411 MW; C5779633037D1223 CRC64;

Query Match 6.6%; Score 211; DB 16; Length 676;
Best Local Similarity 20.9%; Pred. No. 4.5e-05;
Matches 132; Conservative 85; Mismatches 248; Indels 166; Gaps 27;

QY 76 TQOQBEVL---KEQTGLAEAPDNVKTMLR-----SKGYFSSKVSLTEKDGAYTVH 125
Db 115 TPELQEIIRQVIKTQG-----GDTSQLQDRDVAAILLETGLFAS-----ANVNSR 160
QY 126 ITPGPRTKIANGVAILGDLSDGNLAERYRALENWQVGSDFDQDSDWENSKTSVLGA 185
Db 161 TTPSGUNVYQVQPIVRSLSLTGAKALTYVAQPRFQSQIKPISPEGLKQAVAGNQW 220
QY 186 VTRKGYPLAK-----LGNTRAA-----VNPDTATVDLNVVWVDSGRPIAF 224
Db 221 YADNGYNLARVLSIEPNROGILNINVAEGLVSDIKERFYNDGKTIDS-----GNPV-- 273
QY 225 GQFEITGTRYPEQIVSGLARQPGPPYDLDDLLDFQQALEONGHYSGASVQADFRLQG 284
Db 274 -----GGRTKPDFLRQQL-KLQPGVQFOENIVKQDVQOYLRTGLFQSVNVAFAGDATKL 326
QY 285 DRVPVKVSVTEVKRHKLETGIRLDSYGLGGKTYADYVNLFNK----- 327
Db 327 DMI---YELKENCARAINLGGSTNGVDVGLMGTLNYQDQNGIGKNDTLLANVGLSRDQLQ 383
QY 328 -----GY-----IGSVVMDMKY-----ETTLAAGISQPNRYR 355
Db 384 DTKFISPYRDTNSDLGYTVNAPRRREISETFDDEIKLANGDKVREGKVGTLGSLQRPID 443
```

```

QY 356 GNYWTSNVSYNSTTONLEKRAFSGGIWTVRORAGIDARLGAEEFLAEGRKIPGSDIDLGN 415
Db 444 G--WNASLGFNSTRS-----IRDQGNITPTDAQ--GNPLSVSGTGVD--- 483
QY 416 SHATMLTASMKKQLLNVLHPENGHVLDCKIGTTLGTFLSSTALIRTSARAGYF-ETPEN 474
Db 484 -DLTTVSFSATKDQRNDPNINPTQGSV--RVSTEQSVPIGQGNISMNRLKADYSQVVPVN 540
QY 475 ----KKLGTFIIRGOAGTYVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLP 530
Db 541 IFNSQTPQVFPALNVQAGTVL---GNLPPYETFNLGSSNSVRGYDAGNVGSG----- 588
QY 531 ERALLVGSLEYQLPFTRTLSGAVFHDMD-----GDAA-ANFKRMKLKHGSGL-----GVRW 579
Db 589 -RSYVLASAEYRFPPIVPIVGGVLFADFASDLGSGDVTVLGNPAGVRCGKPGSGFGYGAGIRV 647
QY 580 FSPLAPFSDIAYGHSK-KIRWHISLGTFRF 609
Db 648 DSPPLGLIRAD--YGINDQGESRVHLGIGQRF 676

RESULT 33
Q9A711 PRELIMINARY: PRT: 769 AA.
AC Q9A711;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Outer membrane protein.
GN CC1915.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005865; AAK23890.1; -.
DR TIGR; CC1915; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 769 AA; 85938 MW; 7B12A1FEB6226F2B CRC64;

Query Match 6.6%; Score 210.5; DB 16; Length 769;
Best Local Similarity 22.1%; Pred. No. 6e-05;
Matches 145; Conservative 82; Mismatches 247; Indels 183; Gaps 29;

QY 61 DSEIKDMVEEHLPLTIQQOFEVLVDKEQTGLAEAPDN-----VKTMLRSKY 108
Db 187 DNDLRDV-----IVTKESRWYKILTSNDNDYDDRIEYDRQLRKHRYNRGY 232
QY 109 FSKV--SLTE----KDG-AYTVHTPGPRTKIANYGVAILGDLSDGNLAERYRALEN 161
Db 233 EDRFVLISSVAELAPDKNGFAVITYTLEEGPKYFGKITVETELKKL-DGNLAAQI----- 285
QY 162 WQOPV--GSDFDQSDWENSKTSVLGAVTRKGYPLAKLGNTRAVNPDATVDLNVVDSG 219
Db 286 --LPVRTGLYEDERIEQATDALTFAGAAGFAFVDV-RPRVYPNRETKTVDVVFQVREG 342
QY 220 RPIAFGDEEITGTRYPEQIVSGLARQPGTPYDLDLLDFQQALEONGHYSGASVQADF 279
Db 343 PRVYVDRI DIVGNTRTLDYVLRRELEVAEGDAYNRVLVDRSKNMRRRLGFFFEVEIE-DA 401
```


KW Complete proteome.
SQ SEQUENCE 861 AA; 92307 MW; 1C5DFFD394DA5B44 CRC64;

Query Match 6.4%; Score 202.5; DB 16; Length 861;
Best Local Similarity 20.4%; Pred. No. 0.00023;
Matches 131; Conservative 77; Mismatches 254; Indels 179; Gaps 23;

QY 76 TQOQEVLDKEQTGFLEAEAPDNVKTMLRSKGYFSKVSLSLTKDGAYTVHIPTGPRTKIA 135
DB 293 TQIQEDV-----NAIATGYFSN-VRVAPSDPTPLGVRTFVEQANPV 333
QY 136 NGV-----AALGDLSGNLAAYRNALNMWQPVGSDFDQDQSWENKSTVS 182
DB 334 FTGLNIRTPETAEGKERLPOEVVDETGEQYKGKILNRLRELQEGIKTINWYSN-----388
QY 183 LCAVTRKGYPLAK-LGNTRAAVNPOTATVNLVWVDSGRPIAFGDFEITGTORYPEQIYS 241
DB 389 -----QGYDLAQVSGVQVQAGQVTLVIAGVVENIQ-VRFDS-----DEPVG 433
QY 242 GLARFQGPYPYDLILLDDFQQALEQNGHYSGASVQADFRLQ-----284
DB 434 GRTR-----DFIITREMLKPCDVENRNRRAQTDLQRYVSLGLFEDVRLSFNPGSDP 484
QY 285 DRVPKVSVEYKRLKLEGTIRLDSYGLGKIAYDYNNFNKGYIGSVVWDMKYETTL 344
DB 485 TEVINVDVVEGNTGSIAGGGISSSGLFTISYQERNLGGNNQITIGVEAQVGORELLF 544
QY 345 AAGISQPRNYRNYTNSVYNRSTONLEKR-----AFSGGIWYVRDRAGIDA-----393
DB 545 DVSFTDP-----WIGDPFTSTYANLFRRTISLVFDGADSSIRTPNGFDSRVRVT 597
QY 394 -----RLCAEFLAERKTI-----PGSDIDL 413
DB 598 GLGLTFEPIADVDVAPPDWRLSAGGYQNVRIENAGALSPPSAPLNGFNSQPLSFSY 657
QY 414 GNSHATMLTASWKRLNLLNVLHPENHYLDGKIGTTLGTLSTALIRTSARAGY-FTFP 472
DB 658 GVDELFTLSFGASQDNRNNALOPTSGSLV--RFGAEQITPVGTGNIMMTRLRGSYSYIP 715
QY 473 -----ENKLGTFIIRGAGTYVARDNADVPGLMFRSGGASSVRGYELDSIG 520
DB 716 VNWDLTGELVESTQPTVAENVAGTVL-----GDLPEYEAFLGGSNSVRGYQGEGL 771
QY 521 LAGPNSVLPERALLVGSLEYQLPTRTLGSVAFVDMGD-----AAANFKRM-----KL 569
DB 772 ----NG-----RSEFOATAEYRFPITIAAVGGALFVDYGSNLGSQAGVPGFPAIVRGLPGS 822
QY 570 KHGSLGVWFWFPLAPFSDIAY-GHSDKKIRWHISLGTFRP 609
DB 823 GVGYGLGVRIQSPVGPVIRIDLGTGEGESRI--NFGIGKEF 861

RESULT 36
Q9PKF3
ID Q9PKF3 PRELIMINARY; PRT; 792 AA.
AC Q9PKF3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Outer membrane protein, putative.
GN TC0512.

OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;

"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT RT pneumoniae AR39.*;
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002320; AAF39354.1; -.
DR TIGR; TC0512; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 792 AA; 88824 MW; 18387EC35B0F5666 CRC64;

Query Match 6.3%; Score 199; DB 16; Length 792;
Best Local Similarity 22.0%; Pred. No. 0.00034;
Matches 136; Conservative 76; Mismatches 232; Indels 174; Gaps 30;

QY 66 DMVEEHLPLITQOQEVLDKEQTGFLEAEAPDNVKTMLRSKGYFSKVS---LTKDG-- 120
DB 228 DMVDQDLFAIT-----NYFONKGYADAKVTKVEYSTDAKNL 263
QY 121 AYTVAHITPGPRTKIANV---GVAILGDLSGNLAEYRNALNMWQPVGSD---PDQDS 174
DB 264 SLVINVDKGPPLYTLGHVHTEGFTALSKRLDKQL-----VGPNSLYCPEV 310
QY 175 WENSKTSVLGAVTRKGYPLAKLGNTRA-----AVNPDATVLDLVVVDGRPIAFGDFEIT 230
DB 311 WAGAO-KIRNAVARYGV-----NTNVDVSFAVHPPTLPVVDVTVRVSSEGAPYKIGLIKIK 364
QY 231 GTQRYPEQIVSGLARFQPGPYDILLDDFQQALEQNGHYSGAS---VOADFRLQDGR- 286
DB 365 GNTHKHVDVILHETSLFPQDFTDLRLKLEDTETRLRNTGYFKSVSVYTVRSQDLPDLSNLL 424
QY 287 ---VPVKVSVTEV-----KRHKLETG-----IRLDS 310
DB 425 YRDVFIEVKETETGNLGLFGSSIDHLFGGAEISESNFDLFGARHLFKKPKSLRGGE 484
QY 311 Y-----GLGKTIAYDYNLFNKGYIGSVVW-----DMKYETTLAAGISQPRNVRGNYWTS 361
DB 485 YLFKANLGDKVT-DYTVKWTKPHFLNTPWILGVELDK---SINKALSK-----DYSVD 534
QY 362 NVYSNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATML 421
DB 535 TYGCNISTTYIINDK-LKYGMYY-----RGSQTSLSLRKKTENPNKPCPDLD---SKGFPV 586
QY 422 TASWKRQLLNVLHPEN-----GHYLDGKIGTTLGTLSTALIRTSARAGYFFTP 472
DB 587 SAAGNLVYDSIDNPRKPTMGIRSSLNFELSGLGTYQTKLTA-----SGSIYRL 637
QY 473 ENKLGTFIIRGQA-----GYTVARDNADVPGLMFRSGGASSVRGYELDSIG-----LA 522
DB 638 LTKK-GVLKIRGEAKFIKPGFTTAAQ---GIPVSEFFLGGESTVRGYKPFIIIGPKFSPT 693
QY 523 GPNGSVLPERALLVGSLEYQLPTRTLGSVAFVDMGDAAANFKRMKLKH---GSLGVVR- 578
DB 694 EPQGGI--SSLLTTEEFQPLISQPSINAFVFLDSFGITGEETYIRLKNLCSAGFGLAF 751
QY 579 -----WFSPLAP 585
DB 752 DMNANNVPIMLGWGWPPRP 769

RESULT 37
Q9PEI2
ID Q9PEI2 PRELIMINARY; PRT; 784 AA.
AC Q9PEI2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane antigen.
GN XF1046.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
OX NCBI_TaxID=2371;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
RL Nature 406:151-159(2000).
DR EMBL: AE003941; AAF83856.1;
DR InterPro: IPR00184; Bac_surfaG_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 784 AA; 87991 MW; 719331DD881E8192 CRC64;

Query Match 6.2%; Score 197.5; DB 16; Length 784;
Best Local Similarity 20.5%; Pred. No. 0.00042;
Matches 136; Conservative 105; Mismatches 251; Indels 131; Gaps 28;

Qy 36 LFKNKPDESIVKPKPPVR-----IDQDSEIKDMV-----EHLPLITQOQEVLDKE 86
Db 164 LDRNRVDVTIAIKEGKAIAKIRHINLIGTEKFNKNDVMSAWESKEHNWASWYRDDQYSKE 223
Qy 87 Q-TGLAEAPDNVKTMLRSKYF-----SSKVSIT-EKDGAY-TVHTTPGPKIANVG 138
Db 224 KLSGDL-----EKUNSWLDRGYVDNFIDTSQVSPSEKHNMFITAGVTGDDQYKISSIK 278
Qy 139 VAILGDILSDGNLAEYRNALENNMQP-VGSDFDQDSDWSENKTSVLGAVTRKGYPLAKIG 197
Db 279 VT-----GNTV-LPQEKIEKLVIPKTDGIFSRVLLEYSSAALINTLSNIGYAFSKV- 328
Qy 198 NTRAAVNDPTAVDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTYDLDLL 257
Db 329 NPIPTANRADRTAVANLHVIPGPRVTVRQILFKGNTRTSDVLRMRQFNSWYQAAI 388
Qy 258 LDFQALQNGHYSGASVQADFRLQGRVPVKVSVTEVKRHKLETGIRLSEYCLGKI 317
Db 389 DRSKIRLQRLGYFAVDVESTPVGSDNDQVDIVTVYVTRKTTSGSFQVGLGYSKTYGVTTSV 448
Qy 318 AYDYNNLFNKGYSIGSVWMDKYEITTLAAGISQPNRYGNWTSN-----VSYS 365
Db 449 QLSQNNFLGSGNRVSDASRSYQDRYSFSYTNP-----FFTDNGVSLGYNLAYOKLDY 502
Qy 366 NRSTQNLEKRAFSG-----GI-----WYVRDAGIDARLGAELAEGRKIPGSD-- 410
Db 503 SDFNAAQYNSKRMSCQITIFGIPITENDTVSNWI-----CADSNQITTF-----PGSNPK 551
Qy 411 --IDLGNS-----HATMLTASWKNOLLNNVLPENHGY-----LDG-----KIGT 448
Db 552 AIIDYIDAVGORTFRWTELGWARDTRNDYFMPNLGMQRIGAEVTLPGSTIKYIKYIN 611

Qy 449 TLGTF--LSSTALIRTSARAGYFFTPENKKGITFIIRQAGYTVARDNADVPSSL-----M 502
Db 612 QISKYWPILPALVLENTLEVGY-----GDDYKGKSHRILPDGTVA-----TASGLPFFEN 661
Qy 503 PRSGASSVRGYELDSIGLAGP-----NGSVLPERALLVGSLE-----YQLPFTRTL 550
Db 662 FYAGTNSVRGFRNTLGPSEVSEVTAALYNQOGLAGSEKTVGSTEMYFPLKDFSPSARISA 721
Qy 551 GAVFHDMDAAANFKMKKHKHKGSLGVWVFPPLAP-----FSFDIAYGHSDKKIRWHISIG 606
Db 722 FLDFGNVFNNGVNNFKANELRASSGVALLWRAPIGPISISYAPPIKKNENDEIERLQFTFG 781
Qy 607 TRF 609
Db 782 GQF 784

RESULT 38
Q8YHHO
ID Q8YHHO PRELIMINARY; PRT; 781 AA.
AC Q8YHHO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Outer membrane protein.
GN BMEI0830.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RA MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapratl V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009524; AAL52011.1;
DR InterPro: IPR00184; Bac_surfaG_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 781 AA; 85918 MW; E3C7385CCB11DB3A CRC64;

Query Match 6.1%; Score 194.5; DB 16; Length 781;
Best Local Similarity 19.5%; Pred. No. 0.00064;
Matches 141; Conservative 115; Mismatches 284; Indels 183; Gaps 30;

Qy 26 LSENKAAGFALP-----KNKSPD-TESVKLKPFPVRIDTQDSEIKDMVEEHLPLITQOQE 80
Db 103 VKERSVNVNVLFGQNKIKDPDLARAVOLKPRAPDMATMEAD-----K 146
Qy 81 EVLDKEQTFGLAEAPDNVKTMLRSKYFSSKYSITEKDGYTVHTPGPRTKIANVGVA 140
Db 147 EAIKAANYSHIGRSDATNARTVDLGQG-----RVNV-----VYEINGSRKIANIEFV 195
Qy 141 -----ILGDILS-----DGNLAEYRNALENNMQPVGSDFDQDSDWSENKTSVLGAVTRK 189
Db 196 GNOAFSGRRRLRDVISTKRSNPLSWLTRNDV-----YDEGRLOADEETLRRFYNR 245
Qy 190 GYPIAKLGNTRAAVNDPTAVDLNVVDSGRPIAFGDFEITGT-QRYPEQIVSGLARFQ 248
Db 246 GYADPFRVLSSNAVLPDSTNEVYITITVDEGPRYTFGDVSVESVTDGVDVTDALDRLVKRT 305
Qy 249 GTPYDL-----DLLDFOQALQNGHYSGASVQADFRLQGRVPVKVSVTEVKRHKLE- 302
Db 306 GKPYSAKIEISVLSVTSVAVSGS-YAFKAVEPRGDRNFENHTISVYVSDGQPRAYTOR 364
Qy 303 ---TGIRLDSYVGLGGKIAYDYNNLFNKGYSIGSV---VMDMKYET---TLAAGISQPR- 352

```
Db 365 IEIRNDKTRDYVIRREFDLNEGDAFNQVMQAKRRRLVLELDFQTVNISTAPG-SEPQD 423
Qy 353 -----NYRGNYWTSNVSYNRSTQNLKRAFPSCGGIWIYVRDRAG----- 390
Db 424 VILVVDVVEKSTGEFSIGGGYTTGGESPGAQVAAITERNFLRGQYIRISAGAGQDDMR 483
Qy 391 -----IDARLGAELAEGRK-----IPGSDI----- 411
Db 484 NYGLSFTPEYFLGYRLSAGFDVFRSRVYRNDYDVEQTGGTIRFGLPITDNFSAGIAYSL 543
Qy 412 -----DLGNHATML-----TASKR-----QLNVLHPENGHYLDGKIG 447
Db 544 VOEKYDLFRGDAENTYAPALLAEANSPWLRSSVSYSLTYSIDDKNPHDGLY--GRFI 601
Qy 448 TTLGTFLSALTIRTSARAGYEFTPENKLGTFITRGOAGYTVARNADVPGLFRSGG 507
Db 602 QEFAGLGDAKVKYTKFKNNYQTLISOEADIVGLLGVGAGYIHERFGDCVRIFDLFKN-S 660
Qy 508 ASSVRGYELDSIG--LAGPNS--VLPERALLVGSLEYQ-----LPFTRLTSGAVF----- 554
Db 661 SDIIRGFENGIPGYODAKNGRYNMGGTTYFSGTAEOQFPMPLPESLGVRGAFFADAA 720
Qy 555 ----HDMGDAANFRKMLKHGSGGLGVRFSPFLAPSFEDIAY-----GHSDKKIRWHISLG 606
Db 721 TLYGNDTPDISGDDK--KLRSVGVSLMAWSPFGLREDYAFPPVAKADTKVQNFNFGVS 778
Qy 607 TRF 609
Db 779 TRF 781

RESULT 39
Q93PM2 PRELIMINARY; PRT; 793 AA.
AC Q93PM2
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Integral outer membrane protein.
GN D15
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
OC Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21295095; PubMed=11401984;
RX Thomas K.L., LeBuc I., Olsen B., Thomas C.E., Cameron D.W., Elkins C.;
RT "Cloning, Overexpression, Purification, and Immunobiology of an 85-
RT Kilodalton Outer Membrane Protein from Haemophilus ducreyi.";
RL Infect. Immun. 69:4438-4446(2001).
DR EMBL: AF329831; AAK70345.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR SEQUENCE 793 AA; 88661 MW; E4E9D862A8F9903F CRC64;

Query Match 6.1%; Score 193.5; DB 2; Length 793;
Best Local Similarity 19.2%; Pred. No. 0.00076;
Matches 150; Conservative 104; Mismatches 278; Indels 249; Gaps 30;

Qy 27 SENKAAGFALFNKNSPDTSVSKLPKFPVR-----IDQDSEIKOMVEEHLPLITQOO 79
Db 64 SONR-----FENVSAAREGQTLVTKVAERPLINLTIKGNNAPKNALEQNKANLIVA 117
Qy 80 EEVLDEQTCFLAEAPDNVNTMLR-----SKGYESSKVSLETKDCAYT----- 123
Db 118 GEVYDKAKLEAPKQALVDHYHTMGVQADIQITTPPNNSINVELNITEGVIAYVKKN 177
Qy 124 -----VHITPGP-----RFKIANVGVAILDGLSDGNLA----- 152
Db 178 FEGNNAFSYDELKEIKRPNAPWNWIFESSKFOQOEYKNDIEILRDFYMDHGYAKFTLK 237
```

```
Qy 153 -----EYRNALE---NWOOPVGSDF-----DODSWENSKTSVL-----GAVTRK-- 189
Db 238 OTDVKFNENKTEVDLTKYKINEGSQYNISEMRIIGDTQKLDNELNLLTHFKAGQLFRKTE 297
Qy 190 -----GYPLAKLGNTRAAVNPTATVDLNVVVDSCRPIAFGDFEITGTQR 234
Db 298 LSIIEEQIKQILGDRGYSKAV-DLYPKFNEEDHTVQINFIADVGRRIYVRKIRFEGNDV 356
Qy 235 YPEQIVSGLARFPQTPYDLDLLDFQOALEQNGHYSGASVOADFRLQGDRAVPVKVSYT 294
Db 357 TADSTLRLREMROQEGANLSTSAVSLAKSLRTEGYETVEVEMSPVTKNTDQDVIDIYKIK 416
Qy 295 EVYKRHLKETGIRLDSYGLGGKIAYDYNLFNKGYIGSVVWMDMYETTLAAGISOPR-- 352
Db 417 ERNTGSINFGV---GYGSGLSYN-----AGITQDNFL 447
Qy 353 -----NYRGNTWTSNVSYNRSTTQ---NLEKRAFSGGIWYVRDRAGIDARLGAELAE 403
Db 448 GMGSLGLNGSRNTDSTNVNLSYTEPYFTKDCVSLGGNIFY-----EDYDNSARKASAAYK 503
Qy 404 RKIPGSDIDLQ---NSHATMLTASWKROLLNNV-----LHPENCHYLD 443
Db 504 RYTYGASGTGLGFPVDENNSYILGLGYTHDKLRNVEREYTRKYYVNSMKPPIPNQSHY-- 561
Qy 444 GKIGTTLGTFLSALTIRTSARAGYEFT-----PENK--KLGT----- 479
Db 562 DRIQS--ADFDLISFGWNYNNLRGYFPTAGSSANISGKLTLPGSDNKYQVGTNFSGYIP 619
Qy 480 -----FIIRGOAGYTVARNADVPGLMFRSGGASVVRGYELDSI-----GL 521
Db 620 LNSEHKWVIATKGLAYTNSFGGKEVPFYQLYSAGMGSLRGFAGGSGIGPKAIYYREDGF 679
Qy 522 AGPNGSVLPERALLVGSLEYQLP-----FTRLTSGAVF----- 554
Db 680 KAPSDQVIGGNAMVNASLELIIPAPFISDKYQHNVRTSVFVDAATVWNTKWKOSKADYPN 739
Qy 555 -HDMGDAANFRKMLKHGSGGLGVRFSPFLAPSFEDIA-----YHSDKKIRWHISLGR 608
Db 740 LPDFGD---YKRVRA--SAGIALQWQSPIGPLSFYAKPIKYYA-GDEIEQFQFTVGST 792
Qy 609 F 609
Db 793 F 793

RESULT 40
Q92Q48 PRELIMINARY; PRT; 776 AA.
ID Q92Q48
AC Q92Q48;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative outer membrane transmembrane protein.
GN OMP OR R01502 OR SMC02094.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Gallibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591787; CAC46081.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
```


KW Complete proteome.		SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;	
Query Match		6.0%; Score 189.5; DB 16; Length 776;	
Best Local Similarity		21.7%; Pred No. 0.0013;	
Matches 130; Conservative		87; Mismatches 245; Indels 137; Gaps 30;	
Qy	80	EEVLDKKQGTG---FLAEE---APDNVKT---MLR-----SKGY-----FSSKVSLSLTERKGA	121
Db	206	QSVIATKESGIFSLTRKDVNPKLRADELLRQFYNNRGYADFQVVSSEAALNEATNE	265
Qy	122	YTVHIT--PGPRTKIANGVAILGDLISDGNLAIEYYNLENWQOPVGSDFDQDSWENSK	179
Db	266	YTVITIEEGPRYDFGPVNI-----ESTVEGIDAEELRGLVQSRE--GTVYKAKDIQSTM	318
Qy	180	TSVLGAVTRKGYPLAKL---GNTRAAVNPDTATVDLNNVVDSDGRPIAFGDFEITGTORYP	236
Db	319	SEISKRVASEGYPARVTPGRNRLA---NHTIADVLYDQGERAYVERIEIRGNTRT	374
Qy	237	EOIVSGLARFQPGTPYDLDLLDFQQALEONGHYSGASVQADFRLQGDGRVPVKVSVTEV	296
Db	375	DYVIRREFDVGEGDAFNOEMVARAKRRLAALGYFSSVNIQTQ--PGSAADRVVIVVDVQDQ	433
Qy	297	KRKLETGIRLDSYGLGG----KIADYNNLFNKG-YIGSVVW----DMOKYETT-----	343
Db	434	STGSF--GIGAGYSAGDGGGFLVEASIEEKNFLGRGOYIRLAAGKGEDSQTNYVSFTEPY	491
Qy	344	-----LAAGISQPRNYRGNYWTSMVSYNRSSTQNLKRAFSGGIWYVDRAGIDARLGAE	398
Db	492	FLGYRLAAGFDLFXN-ENDFDDNNYSN-----DQGF-----LRVTAPITENLSTT	537
Qy	399	FLAEGRKIP--GSDIDLGNSHATMLTAS-WKR-----QLLNNVLHPENGHYLDGKI	446
Db	538	LRVNYTELYFGDRDELSSPYDRVIDGSPWTRSSISQITVNTLDDAQLPHEG-----	590
Qy	447	GTTGLTFLSSTALRTSARAGYFTTPENKKLGTFIIRGQA-GYTVARNADV-----PS	499
Db	591	-----ILASV---TQEFAGLGCTSDFYKL-----TGKAKWYTLHDEADIIGLSGSA	635
Qy	500	GLMFRSGG-----ASSVRGYELDSTIGLAGPNGSVLPERALLVGSLEYQ-----L	543
Db	636	GHLFETSGSLEVFDQFOLNSNDIIGFERNGLGPRMNGDALGGTTYFTASAEATFPLPGL	695
Qy	544	PFTRTLSCAVPHDMGDAANFKRM-----KLKHGSGLGVRWFSPILAPFSFDIA	591
Db	696	PRDSGFRGALFVDAGTLYGNDVEIGPGESVRGDNASLRASVGVSLINWASPGPLRVDA	754

Search completed: November 9, 2002, 01:20:14
Job time : 39 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:14:29 ; Search time 17 Seconds
(without alignments)
1485.828 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 NM1KPTALLPALFFPHAY.....IAYGSHDKKIRWHISLIGTRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	379.5	12.0	577	1 YTFM_ECOLI	P39320 escherichia
2	374	11.8	578	1 YTFM_HAEIN	P44038 haemophilus
3	217	6.8	810	1 UP05_ECOLI	P39170 escherichia
4	213.5	6.7	793	1 D153_HAEIN	O32629 haemophilus
5	209.5	6.6	795	1 D152_HAEIN	P44935 haemophilus
6	209.5	6.6	797	1 D151_HAEIN	P46024 haemophilus
7	138	4.3	835	1 FASD_ECOLI	P46000 escherichia
8	133	4.2	779	1 K6PF_HUMAN	P08237 homo sapien
9	131	4.1	781	1 K6PF_CANFA	P52784 canis famil
10	128	4.0	779	1 K6PF_RABIT	P00511 oryctolagus
11	124	3.9	881	1 YEBT_HAEIN	P44288 haemophilus
12	121.5	3.8	617	1 Y237_BUCAL	P57331 buchnera ap
13	120	3.8	779	1 K6PF_RAT	P47858 rattus norv
14	119	3.8	779	1 K6PF_MOUSE	P47857 mus musculu
15	118.5	3.7	815	1 FPVA_PSEAE	P48632 pseudomonas
16	116.5	3.7	1300	1 L20K_RICRI	P14914 rickettsia
17	116.5	3.7	1654	1 OMPE_RICRI	O53047 r outer mem
18	116	3.7	922	1 MPPL_CHLPPN	Q29955 chlamydia p
19	114.5	3.6	3305	1 APLP_MANSE	Q25490 manduca sex
20	113.5	3.6	1018	1 VGNM_BPMV	P23009 bean-pod mo
21	113.5	3.6	1577	1 HLVA_PROMI	P16466 proteus mir
22	113	3.6	838	1 YRAJ_ECOLI	P42915 escherichia
23	111.5	3.5	595	1 VPL_BPCHP	P19192 bacterioph
24	111	3.5	574	1 TACY_STRCB	Q53957 streptococc
25	111	3.5	976	1 FIBP_ADEB3	Q03553 bovine aden
26	110.5	3.5	691	1 YHJG_ECOLI	P37645 escherichia
27	110	3.5	828	1 MRKC_KLEPN	P21647 klebsiella
28	110	3.5	1423	1 FRUA_STRMU	Q03174 streptococ
29	109	3.4	571	1 TACY_STRPY	P21131 streptococ
30	109	3.4	584	1 FHAC_BORPE	P35077 bordetella
31	109	3.4	1076	1 CARB_VIBCH	O9kph9 vibrio chol
32	108.5	3.4	841	1 MYFC_YEREN	P33408 yersinia en
33	108.5	3.4	1375	1 GTFC_STRMU	P13470 streptococ

NO

34 108 3.4 517 1 HSF2_MOUSE
35 108 3.4 574 1 FLA3_CAMJE
36 107.5 3.4 479 1 PRTC_ERWCH
37 107 3.4 587 1 P1CP_PSESR
38 107 3.4 791 1 K6PP_RABIT
39 107 3.4 859 1 AFAC_ECOLI
40 106.5 3.4 356 1 E2B1_PYRAB
41 106.5 3.4 1320 1 PUTA_ECOLI
42 106.5 3.4 1655 1 OMPB_RICCN
43 106 3.3 3034 1 CLB1_MOUSE
44 105.5 3.3 812 1 FRED_ECOLI
45 105 3.3 574 1 FLB3_CAMJE

P38533 mus musculu
Q46113 campylobact
P16317 erwinia chr
P42790 pseudomonas
P47859 oryctolagus
P53517 escherichia
Q9uz16 pyrococcus
P09546 escherichia
Q9kka3 r outer mem
Q35161 mus musculu
P06970 escherichia
Q46114 campylobact

ALIGNMENTS

RESULT 1
YTFM_ECOLI
ID YTFM_ECOLI STANDARD; PRT; 577 AA.
AC P19320:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ytfm precursor.
GN YTFM OR B4220 OR Z5831 OR ECS5198.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / WC1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes";
RL Nucleic Acids Res. 23:2105-2119(1995).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / BDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Davis N.W., Lim A., Dimalanta E.T., Miller L.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI0698.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC

```

MEDLINE=95350630; PubMed=7542800;
RA Fleischmann A.R., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatava A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merriick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.B., Frichman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenza
RT Rd."
RL Science 269:496-512(1995).
RL [2]
RN IDENTIFICATION BY MASS SPECTROMETRY.
RP MEDLINE=20137488; PubMed=10675023;
RX Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RX Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae."
RT Electrophoresis 21:411-429(2000).
RL -1- SIMILARITY: STRONG, TO E.COLI YTFM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabor
CC between the Swiss Institute of Bioinformatics and the EMBL outstat
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commu
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32752; AAC22357.1; -.
CC TIGR; HI0698; -.
DR InterPro: IPR000184; Bac_surfag_D15.
DR Pfam: PF01103; Bac_surface_Ag; I.
KW Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 578 PROTEIN HI0698.
FT SEQUENCE 578 AA; 65812 MW; 34F9AC189C505876 CRC64;
SQ -----
Query Match 11.8%; Score 374; DB 1; Length 578;
Best Local Similarity 22.9%; Pred. No. 7.4e-17;
Matches 153; Conservative 93; Mismatches 263; Indels 158; Gaps
QY 3 IKPTALLPALFFFPFHAYAPADLSENKAAGFALFNKSPDTSVKLPKFPVPRIDTQDS 62
Db 6 LKLTALFL-ALSCFPAPAEQTVDIEVQGI RFAVRNTD-----LNVLNLIK 51
QY 63 EIKDMVEEHLPLITQOEEVLDKEQTGFLEAEAPADNVKTMLSKGYFSKVSLSUTEKD--- 119
Db 52 EEMDGSERYOHLVTKA-----VDRGLRVFGYESSVFRKQROG 91
QY 120 --GAYTVHITPGPRTKIANGVAILGDILLSDGNLAEYVYRNALENMQPVGSDFDQDSWEN 177
Db 92 KROLLTAHVTTPGEPTKIAGTDVQIEGAQAQENF-----NALRKNLPKQGLVVEIQTYYD 146
QY 178 SKTSVLGAVTRKGYPLAKLGNTRAAYNPDTATVDLNVVVDSCRPIAFGDFEITGYQYPE 237
Db 147 YKTAISRALNRGYFDGNGFKISRLSIPETHQAWRMFLDSGVRVHYGNITFSHSQ-IRD 205
QY 238 QIVSGILARQPGPYDLIDLLDLPQQALEONGHYSGASVQADFRLQGDVRPVKVSYTEVK 297
Db 206 DYUNNLINTKSGDPIYLMNNLSLDTSDFFPSNWMFSSVLVQPNVNH-KSKTVDDVEILYPRK 264
QY 298 RHKLETGIRLDSYGLGGKIAY----- 319
Db 265 KNAMELGVGFTDGGVHGQIGWTKPWINSRHSLSRNLVLSAPKQTL EATYRMLPKNPL 324
QY 320 DYNLNLKNGYIGSVWMDKQYETTLAAGTSQPRNYRGNTWTS-----NVSYNRST 369
Db 325 NYIYDFAVGWEGEKENDTNRVLTLISA-----LRYNNAHGQWYFGGLRMRYDSPT 375
QY 370 TONLEKRAF----SGGIWVVRDRAGIDARLGAEEFLAEGRKIPGSDIDILGNHATMLTASW 425

```

```

Db 376 QADITDKLLLYPTVGTFRRLRGLSGFATWG-----DVQKI-----TFDLS-----416
QY 426 KROLLNNVLPHNGHYLDGKIGTTLGFLSSTALIRTSARAGYFFT-PENKKGTLGIIRG 484
Db 417 KR-----IWLSSEFIKQVASSAVRYAENHRV---VARA 449
QY 485 QAGTYVARDNADVPGLMFRSGASSVRGYELDSIGLAGPNSVLPERALLVGSLEYQLP 544
Db 450 EIGLYHTKGLEKIPPTLRFPAGDGRSVRGYKKIAPKNRNGKLVGSRLLTTSLEVQYQ 509
QY 545 FRTLSGAVPHDGDAAAFKMKLKHGSLGVRWFSPLAFSFDIAYGHSD-----KKIR 600
Db 510 VYPNWAATFADSLAADNYTAKELRVGTGVGRWASPVGAIFEDIATPIRDKNKNIQ 569
QY 601 WHISLGT 607
Db 570 FYIQLGT 576

RESULT 3
ID UP05_ECOLI STANDARD; PRT; 810 AA.
AC P39170; P39181; P77465;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page spots M62/M63/O3/O9/T35 precursor.
GN VAEF OR B0177 OR 20188 OR ECS0179.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97443975; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).

```

```

RN [5]
RP SEQUENCE OF 21-32 AND 351-362.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12."
RL Electrophoresis 18:1259-1313(1997).
CC 1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC 1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000127; AAC73288.1;
DR EMBL; U70214; AAB08606.1;
DR EMBL; AE005193; AAG54479.1;
DR EMBL; AF002550; BAB3602.1;
DR SWISS-2DPAGE; P39170; COLI.
DR EcoGene; EG12676; yaeT.
DR InterPro; IPR000184; Bac_surfac_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 810
SQ SEQUENCE 810 AA; 90552 MW; DDCE4C6D341664EB CRC64;

Query Match 6.8%; Score 217; DB 1; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.6e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;

QY 97 DNVKTLMLRSKGYF-----SSKSLIT-EKCGAY-TVHITPGPRTKIANVGAILGDILSDG 149
Db 224 ETLSRYLDGRYARFNIDSTQVSLTPDKGIYVTVNITGDKYKLSGVSVS-----C 275

QY 150 NLAEYRNALENMOQPVGSDFDQDSDWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTAT 209
Db 276 NLAGHSAEIEQLTKIEPGELYNGTKVTMEDDIKILGRYGYAVPRV-QSMPEINDADKT 334

QY 210 VDLNVVDSRPTAFGDFEITGTPQVPEQIVSGLARFOPGTPDLDLLDFQALEONGH 269
Db 335 VKLRNVNDAGNRFYVRKIRFEGNDTSKDAVLRREMQMEGAWLGSLDVGQKERLNRLGF 394

QY 270 YSGASVQADFRLQG--DRVPEVKVSVTEVKRHKLETCGIRLDSVGLGGKTAIDYNNFNK 327
Db 395 FE--TVDTDTQRPVGPDPQVDVYVKVKERTGSGFNFGIGYGVTSQVSGVQAGVQDNLGT 452

QY 328 GYGSVVMDMDKYETTLLAAGISQPRNRYGNWT-----YFTVDGVSGLGRLFYNDFOADDALSDYTN 362
Db 453 GYAVGINGTKNDYQVLAELSVTNP-----YFTVDGVSGLGRLFYNDFOADDALSDYTN 506

QY 363 VSNRSTQNL---EKRAFSGGTYWVDR-AGTDARLCA-EFLAEGRKIPGSDIDLGNH 417
Db 507 KSYGTDVTLGFPINENSLRAGLYVHNSLSNQPMQVAMWRYLYSMGEHP-STSDQDNSE 565

QY 418 AT---MLTASWKRQLLNVLHPENGH--YLDGKI-----GTTFLFLSSTALIRT 462
Db 566 KTDFTFNYGWTYNNKLDGRGYFTDGSRVNLTGKVTIPGSDNEYKVTLDT----- 615

QY 463 SARAGYFFTPENKKGTLFIIRGQA--GYTVARDNADVPGLMFRSGASSVRGYELDSIG 520
Db 616 ---ATVYPIDDHK---WVVLGRTRWGYDGLGKKNPFYENFYAGSGTSVTRGFSNTI- 668

QY 521 LAGPNSGVLPER-----ALLVGSLEYQLP----- 544;
Db 669 --GPKAVYFFHQASNYDPDYECATQDGAKDLCKSDDAGVGNAMAVASLEFITPTPFIS 726

```

```

QY 545 --FTRTSLGAVPHDMGDA-AANFKRMK-----LKHGSLGVRWFSPPLAPFSFD 589
Db 727 DKYANSVRTSFFWDMGTWYNDWSDSYSGYPDYSNIRMSAIALQWMSPLGPLVES 786
QY 590 IAYG----HSDKIRKWHISLG 606
Db 787 YAPPEKKYDGDKAQEQFNIG 807

RESULT 4
ID D153_HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL: U60834; AAB61977.1;
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFD82036801A14 CRC64;

Query Match 6.7%; Score 213.5; DB 1; Length 793;
Best Local Similarity 21.0%; Pred. No. 2.7e-06;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;

QY 105 SKGYFSSKVSUT-----EKDGA-YTHVHTPGPRTK-----IANVG--VAILGDILSDG 149
Db 228 NNGYAKAQITKTVQNLNDEKTKVNVTVIDNNEGLQYDLRSARIIGNLGMASAELEPLLSAL 287
QY 150 NLAERYNALENMQQVCSDFDQSDWENSKTSVLGAVTRKGYPLAKLGNTRAAYNPD--- 206
Db 288 HLNTDFRS-----DIADVENAIKAKUG-----ERGY-----GNTTVNSVPDPDD 327
QY 207 -TATVDLNVVDSGRPTAFGDFEITGTQRYPEQIVSGIARFPQPYDLDLLDFOQALE 265
Db 328 ANKTLAITFVVDAGRRLTVRLREEGNTVSADSLRQEMRQOEGTWNLSQLVELGKRLD 387
QY 266 QNGHYSASVQADFRLQG--DRVPVKVSVTEVRHKLEGTGIRLDSYGLGKGIAYDYN 323
Db 388 RTGFFE--TVENRIDPINGSNDEYDVVYVKVKNRTGTSINFGYGTSGISYQTSIKQDN 445
QY 324 LFNKGYIGSVVMDKYETTLAAGISOPR-----NYRGNVWTSNVSNYS--TTONLEKR 376

```

```

Db 446 FLGTGAAVSIAGTKNDYCTSVNLGYTEPYFTKDGVSGLGNFFENYDNYSKSDTSSNYKRT 505
QY 377 AFSGGI---WVRDRAGIDARLG-----AEFLAP-----GRKIPGSDIDLG 414
Db 506 TYGSNVTGLFPVNNENSYVGLGHTYKNISNFALEYNRNLYIQSMKFKNGIKTNDFD-- 563
QY 415 NSHATMLTASWKROLNVLNHPENGHYLDGKIGTGTGLTFLSTALIRTSARAGYFFFTPEN 474
Db 564 -----FSFGWYNLSNRGYEPTKG--VKASLGGRVITPGSDNKYKLSADVQGFYPLDR 615
QY 475 KKLGTFLIRGOAGYTVARDNADVPGLMFRSGGASSVRCGYELDSIGLAGPN----- 525
Db 616 DHRWVVSASAGYANGFCNGKRLPFYQTYTAGGIGSLRGFAYGSI---GPNALIAEHGNG 672
QY 526 -----GSVLPERALLVGSLEVQLP-----FTRT-----L 549
Db 673 TFKISSDVIGGNATITASAEILVPTPFVSKDSQNTVTSLEVDAAASVWNTKWSKDKNGL 732
QY 550 SGAVPHDMGDAANFKRMKHKHSGSLGVRWFSPPLAPFSFDIA-----YGHSDKKIRWHIS 604
Db 733 ESKVLKDLPLDYG---KSSRIRASTGVGFQWQSPIGPLVFSYAKPIKKYENDDVE-QFOFS 788
QY 605 LGTRF 609
Db 789 IGGSF 793

RESULT 5
ID D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Uddelman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.W., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32773; AAC22575.1;
DR TIGR: HI0917;
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal; Complete proteome.

```

```
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA: 87478 MW: 85691FC22E6ED44 CRC64;

Query Match 6.6%; Score 209.5; DB 1; Length 795;
Best Local Similarity 20.0%; Pred. No. 4.9e-06;
Matches 130; Conservative 103; Mismatches 27; Indels 141; Gaps 24;

QY 61 DSKDKMVEEHLPLITQQOEEVLDKQGTGLAEAEADPNVKTMLRSGYFSSKVSLT--- 116
DB 184 ESVSSSTLQEQMELQPDSSWMLKGNKFEQAQPEKDIQSDIRDYLLNGYAKAQITKTDVOL 243
QY 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAILGDILSDGNLAAYRNALENWQOP 165
DB 244 NDEKTKVNTIDVNEGLQYDLRSARTIGNLGMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPD---TATVDLNNVVDSSGRP 221
DB 297 -----DIADVENAIAKALG---ERGYGSATVNSV-----PDFDANKTLAITLVDAGRR 343
QY 222 IAFGDFEITGTORYPEQIVSGLARFQGTPTDLDLILDFQQALEQNGHYSGASVQADDFR 281
DB 344 LTVRQLRFEQNTVSADSTLRQENRQOEGTWYNSQLVELGKIRLDRGTGFFE--TVENRIDP 401
QY 282 LQO--DRVPVKVSVTEVRRHKLKLETRLDSEYGLGKIAIDYNNLNFKNKGYISGVVWMDK 339
DB 402 INGSNDEVVYKVKERNTGSINFGIGYGTESGISYQASVKQDNFLGTGAASVIACTKND 461
QY 340 YETTLAAGISOPR-----NYRGNWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDMSKSDTSNRYKRTTYGSNVTLGFPVNNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASMKRQLL 430
DB 522 SYVVGHLGHTYKNKISNFALEYNRLYIQSMKFKGNGIKTNDF-----FSGWYNSL 573
QY 431 NNVLHPENGHYLDKIGTTLGTFLSSTALIRTSARAGYFTFPENKKLGTFIIRGQAGTYV 490
DB 574 NRGVFTPKG--VKASLGGRTVTPGSDNKKYKLSADVOGFPYPLDRDLHLVWVVSASAGYAN 631
QY 491 ARDNADVPGLMPSRGASVRYELDSIGLAGPN-----GSQLPERA 533
DB 632 GFGNKKRLPFQYTAGGIGSLRGFAYGSI---GPNAIYAEHNGNGTFFKKISSDVIGGNA 688
QY 534 LLVGSLEYQLP-----FTRTSLGAVFHDMDGAAANE----- 564
DB 689 ITTASAEIYPTFPVSDKSONTVRTSLF---VDRASVWNTKWSKSDGLDNNVLSLPDY 745
QY 565 -KRMKLKHSGLGVRWFSPLAFSFDIA-----YHSDKKIRWHISLGRTRF 609
DB 746 GKSSRIRASTGVGFQSPIGPLVFSYAKPIKKYENDDVE-QFQFSIGGSF 795

RESULT 6
D151_HAEIN STANDARD; PRT: 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loomore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
Haemophilus influenzae.";
```

```

Db 669 NAIATASAEELIVPTFPYSDKSONQVTSFLFVDAASVNTKWKSDKNGLESVLRLEPDYG 748
QY 568 ---KLKHGSGLVGRWFSLAPESFDIA-----YGHSDKKIRWHISLGRF 609
Db 749 KSSRIRASTGVGFQWQPIGLVFSYAKPIKKYENDVDE-QFQFSIGSP 797

RESULT 7
FASD_ECOLI STANDARD; PRT; 835 AA.
AC P46000;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Outer membrane usher protein fasd precursor.
GN FASD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=987;
RX MEDLINE=94148769; PubMed=7906265;
RA Schifferli D.M., Alrutz M.A.;
RT "Permissive linker insertion sites in the outer membrane protein of
RL J. Bacteriol. 176:1099-1110(1994).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE 987P
CC FIMBRIAE SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; L22659; AAA21827.1; -
DR EMBL; U50547; AAB02687.1; -
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 21 POTENTIAL..
FT CHAIN 22 835 OUTER MEMBRANE USHER PROTEIN FASD.
FT DISULFID 810 834 POTENTIAL.
SQ SEQUENCE 835 AA; 92354 MW; 2FEC6E2305274E1 CRC64;

Query Match 4.3%; Score 138; DB 1; Length 835;
Best local Similarity 17.8%; Pred. No. 0.21;
Matches 104; Conservative 87; Mismatches 201; Indels 192; Gaps 22;

QY 56 RIDTQDSEIKDMVEHLPLITQOQEEVLDRKQGTGLAEAPDNVKTMLRSKGYFSKVSLL 115
Db 259 KIYSENDEMLNRLRSYPTV-----NGIASSQAVV 288
QY 116 TEKDGATVV---HITPGPRTKIANGVAILGDLISDGNLAIEYYRNALENMQQPVGSD--F 170
Db 289 TIKGGVVILQKNVPPGPF-----INDESLSGYSGDLVYNIKEADGSEHSF 335
QY 171 QDQSWENSKTSVLGAVTKRGYPLAKLGNTRAAVNPDTATVLDNVVDSGRPIAFGDEFT 230
Db 336 IQ-----PFSTLPEMKREGV-----SGVEISLGHYNNNS 363
QY 231 GTQRYPEQIVSGLARFQGTGYDLDLLDFOALEQN-----GHYSGASVQADFD 280
Db 364 GATQYYNESPLYASWSENGWNTLSETIQSRKYOLLGVGSLTSLGDGFGAVSGDASLS 423
```

```

QY 281 RL-----QCDRVPVKVSVTEVKRHKLETGIRLD-SEYGLGGRKIAYDYVNLFNKGYIGS 332
Db 424 RANKYDKIHQSQSYGLKYS-----KNKVDGTCTVTLATYRSTKDFSEFNDEFSKNSDVQ 478
QY 333 VYWDMDKYETTLAAGISQPRNYRG-----NYWTSN-VSYNRSTTONLEKRAFSGGI 382
Db 479 YVWD-NRLKNRITLSLNQSLDDYGSLSLIASQQNYWTSDYVSRSFSL-----HSFGWMD 532
QY 383 WYVRDRAGIDARLGAELAEGRKIPG--SDIDL-----GNSHATMLTASMKRQLLNVLHP 436
Db 533 IFSTSFSLQKEGDNALRNKKVFGFYSSIPLSKLKKNESYSTLSYVNTKINNQV-- 590
QY 437 ENGHYLDGKIGTTLGTLFSLTSALIRTSARAGYFFTPENKKLGT----FIIRGQAGYTVAR 492
Db 591 RNTATLAGKVPGSMAQYRFSSGWANTQSS-----NKALSVNWDGDLDDLSGLSYTSSG 643
QY 493 DN-----ADVPSCLMFRSGGASSVRGYEL--- 516
Db 644 KNRTIDYSLSGSAILYPWRLAIGSDSYNGAAVVVETEFISGIKVROGGETSLLGTAVTS 703
QY 517 -----DSTGLAGPNGSVLPERALLVGSLEYQL 543
Db 704 MQPYTENRIDLDTONIPDDLFISNASKKIYVEKCAVV-PVKYNL 746

RESULT 8
K6PF_HUMAN STANDARD; PRT; 779 AA.
AC P08237; Q16814; Q16815;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=92009225; PubMed=1833270;
RA Yamasaki T., Nakajima H., Kono N., Hotta K., Yamada K., Imai E.,
RA Kuwajima M., Noguchi T., Tanaka T., Tarui S.;
RT "Structure of the entire human muscle phosphofructokinase-encoding
RT gene: a two-promoter system.";
RL Gene 104:277-282(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=89306675; PubMed=2526045;
RA Sharma P.M., Reddy G.R., Vora S., Babior B.M., McLachlan A.;
RT "Cloning and expression of a human muscle phosphofructokinase cDNA.";
RL Gene 77:177-183(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=88030023; PubMed=2822475;
RA Nakajima H., Noguchi T., Yamasaki T., Kono N., Tanaka T., Tarui S.;
RT "Cloning of human muscle phosphofructokinase cDNA.";
RL FEBS Lett. 223:113-116(1987).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 271-680 FROM N.A. (ISOFORM 2).
RC TISSUE=Muscle;
RX MEDLINE=90264379; PubMed=2140567;
RA Sharma P.M., Reddy G.R., Babior B.M., McLachlan A.;
RT "Alternative splicing of the transcript encoding the human muscle
```


[illegible]

```
QY 385 VRDRAGIDARLGAELAEGRK-----IPGSDIDLGNHSA--TMLT- 422
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 495 I---GGFEAYTGGLEMEGRQFDELCPFFVVPATVNNVPGSDFSVGADTALNTICT 551
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 -----ASWKRQLLNVLHPENGHYLDGKIGTTLG---TFLSSTALIRTSARAGYFT 471
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 552 CDRIKQSAAGTKRRVF-----IETMGCGYGLATMAGLAAGADAAYIFE 596
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 PENKKGITFIIR---GQAGYTVARDNADVPGLMFRSGGASSVRCGYELDSI-GLAGPNG- 526
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 597 E-----PFTIRDLQANVEHLVQRMKTTVKRGLVLRNEKCN--ENYTTDFIPNLYSEBGK 648
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 527 SVLPERALLVGSLEY---QLPFTRTLSGAVFHDMDGDAANFKRMKLKH----- 571
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 649 GIFDSRKKNVLGHMOGGSPTEPDNRFA---TKGAKAMNMWSGCKIKESYRNGRIEANTP 704
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 572 --GSLGVR---WFSPLAPSFSDIAYGHSKDKIRWHISL 605
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 705 DSGCVLGMKRKRALVFPQVAELKDQDTEHRIPEKQWMLK 744
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9

```
K6PF_CANFA STANDARD; PRT; 781 AA.
ID K6PF_CANFA AC
P52784;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM OR M-PFK.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96194913; PubMed=8654960;
RA Smith B.F., Henthorn P.S., Rajpurohit Y., Stedman H., Wolfe J.H.,
RA Patterson D.F., Giger U.;
RT "A cDNA encoding canine muscle-type phosphofructokinase."
RL Gene 168:275-276(1996).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
fructose 1,6-bisphosphate.
CC -1- COFACTOR: MAGNESIUM.
CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
SUBFAMILY.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```
CC EMBL; U25183; AAC48615.1; -.
CC HSSP; P00512; 3PFK.
CC InterPro; IPR000023; Ppfuckinase.
CC Pfam; PF00365; PFK; 2.
CC PRINTS; PR00476; PPFRCCKINASE.
CC PRODOM; PD000707; Ppfuckinase; 2.
CC KINASE; PS00433; PHOSPHOFRUCTOKINASE; 2.
KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INT_MET 0 0 BY SIMILARITY.
REPEAT 1 403 APPROXIMATE.
```

```
FT REPEAT 404 781 APPROXIMATE.
SQ MOD_RES 776 776 PHOSPHORYLATION (BY SIMILARITY).
SEQUENCE 781 AA; 85429 MW; 15CA0B45438462EE CRC64;

Query Match 4.1%; Score 131; DB 1; Length 781;
Best Local Similarity 19.9%; Pred. No. 0.55;
Matches 129; Conservative 73; Mismatches 236; Indels 210; Gaps 31;

QY 81 EVLDKEQTFLAEEAPDNVKTMLRSKGYFSSKVSITKDGAYTVHTTGPRTKIANVGA 140
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 EIVDAITTAOSHORTFVLEVMGRHCYGLALVTSLS--CGADWFIPECPDP----- 235
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 ILGDTLSDCNLAEYYRYNALENMQPGVSDFDQDSMENSKTSVL-----GAVTRKGYPLA- 194
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 236 -----DWEHLCHRLSETTRGSRSLNIIIVAECAIDKNGKPTISE 276
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 -----KLGNTAAVNPDATVDLNVVDSGRPIAFGDFEITGTQYTPQIVSGLARF 246
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 277 EIKELVVKRLGY-----DTRVTVLGHVQRGGTPSAFD--RILGS-RMGVEAVNMALLEG 326
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 QPGTPYDLDDL-----LDFQQALEQNGHYSGASVQADF--RLQG----- 284
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 327 TPDTPACVVSLSGNOAVRLPLMECVQTKDVTKAMNDRKDFDEAMKLRSPMNNVYKL 386
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 -DRVPKVSVTEVKRHK-----LETGIRLDSEYGL--GGK--IAYDYVNLFNK 327
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 387 LAHIRPPVSKTSATMHTVAVMNVGAPAGMNAVRSTVRIGLIOGNRVLVVHDGEGGLAK 446
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 GYISGVMDMKYETTLAAGISQPNRYGNWTSNVSYNRSTTONLEKRAFS----- 379
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 447 QQIEEAGV-----SYVGG-WTGGGSKLGTKRTLPPKKSPEQISANITK 488
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 380 --GGIWRVDRAGIDARLGAELAEGRK-----IPGSDIDLGNH 417
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 FNIQGLIII--GGFEAYTGGLEMEGRKQFDELCPFFVVPATVNNVPGSDFSVGADT 545
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 A--TMLT-----ASWKRQLLNVLHPENGHYLDGKIGTTLG---TFLSSTALIRTS 463
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 546 ALNTICTCDRIKQSAAGTKRRVF-----IETMGCGYGLATMAGLAAG 590
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 ARAGYFFTPENKKGITFIIR---GQAGYTVARDNADVPGLMFRSGGASSVRCGYELDSI- 519
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 591 ADAAYIFE-----PFTIRDLQANVEHLVQRMKTTVKRGLVLRNEKCN--ENYTTDFIF 642
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 520 GLAGPNG-SVLPERALLVGSLEY---QLPFTRTLSGAVFHDMDGDAANFKRMKLKH----- 571
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 643 NLYSEEGKIFDSRKKNVLGHMOGGSPTEPDNRFA---TKGAKAMNMWSGCKIKESYRN 698
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 572 -----GSLGVR---WFSPLAPSFSDIAYGHSKDKIRWHISL 605
: : : : : : : : : : : : : : : : : : : : : : : : : :
DB 699 GRIFANTPDSGCVLGMKRKRALVFPQVAELKDQDTEHRIPEKQWMLK 746
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 10

```
K6PF_RABIT STANDARD; PRT; 779 AA.
ID K6PF_RABIT AC
P00511;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle.
RX MEDLINE=87166033; PubMed=2951385;
RA Lee C.P., Kao M.C., French B.A., Putney S.D., Chang S.H.;
RT "The rabbit muscle phosphofructokinase gene. Implications for protein
```



```

RESULT 13
K6PF_MOUSE
ID K6PF_MOUSE STANDARD; PRT; 779 AA.
AC P47858; Q63736;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM OR PFK-M.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Pancratic Islets;
RX MEDLINE=96350468; PubMed=8764833;
RA Ma Z., Ramanadham S., Turk J., Kempe K., Hu Z., Ladenson J.;
RT *Characterization of expression of phosphofructokinase isoforms in
RT isolated rat pancreatic islets and purified beta cells and cloning
RT and expression of the rat phosphofructokinase-A isoform.*;
RL Biochim. Biophys. Acta 1308:151-163(1996).
RN [2]
RP REVISIONS TO 162-169 AND 493-497.
RA Ma Z.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 7-212 FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Nakajima H.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate -> ADP + D-
CC fructose 1,6-bisphosphate.
CC -1- COFACTOR: MAGNESIUM.
CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U25651; AAC52786.1; -
DR EMBL; D21865; BAA21013.1; -
DR HSSP; P00512; PPK.
DR InterPro; IPR000023; Ppfckinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PHFRCTKINASE.
DR ProDom; PD000707; Ppfckinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 774 774 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 32 32 T -> A (IN REF. 3).
FT CONFLICT 45 45 L -> A (IN REF. 3).
FT CONFLICT 149 149 R -> A (IN REF. 3).
FT CONFLICT 159 159 F -> G (IN REF. 3).
FT CONFLICT 179 179 S -> F (IN REF. 3).
SQ SEQUENCE 779 AA; 85428 MW; F4C92DDE47D53788 CRC64;

Query Match 3.88; Score 120; DB 1; Length 779;
Best Local Similarity 20.3%; Pred. No. 2.8;
Matches 129; Conservative 72; Mismatches 247; Indels 188; Gaps 31;

```

197	QY	GNTRAAVNP	---DTATVDLNVVDSGRPIAFGDFEITGTORYPEQIIVSGLAREFOGTPYD	2533
277	Db	DIKNLVVKRLGYDTRVTVLGHVQGGTSPSAFD	--RILGS-RMGVEAVWALLEGTPDTPAC	3333
254	QY	LDLL	-----LDFQOALEONGHYSGASVOADEFD	2088
334	Db	VVSLSGNOAVRLPLMECVQVTKDVT	KAMDEKREDEATKRLGRSFMNNWEVYKLLAHVRPP	3933
289	QY	VK	-----VSVTEV--KRHKLEIGIRLDSEYGL--GCK--IAYDYNYLPHKGYIGSVVWD	3336
394	Db	VSKGGLHTVAMVNVGAPAGMNAVRSTVRIGLIQGNRVLVHHDGFEGLAKGQIEAGW	4525	
337	QY	MDKYETTLLAAGISOPRNYRGNWTSNVSYNKRSTQNLEKR	--AFSGGIWYVRDR-----	3688
453	Db	-----SYVGG-WTGCGSKLGTERTLPKANLQGISANTIKFNOGLVII		4955
389	QY	AGIDARLIGABFLAEGRK	-----IPGSDIDLGNSHA--TMLT-----	4222
496	Db	GGFEAYTGGLEMEGRKQFDELICIPFVIVATVSNVNVPGSDFSIGADTALNTICTCDRI	5555	
423	QY	-----ASHKROLLNNVLPHENGHYLDKGKIGTILG	---TFLSSTALIRTSARAGYFFTPENK	4755
556	Db	KQSAAGTKRRVF	-----IETMGYCYGLATMAGLAAGADAAYIPBE--	5975
476	QY	KLGTFIIRG	---OAGYTVARDNADVPSGLMFRSGGASSVRGYELDSI-GLAGPNG-SVLP	5303
598	Db	---PFTIRDLQVNVHVLVOKMKTYKRGVLNRKCN	--ENYTTDFINLYSEGGKIFD	5733
531	QY	ERALLVGSLEY	---QLPFTRTLGSVHFHMDGDAANFKRMKLKH-----GS	5733
653	Db	SRKNVLGHMGQGGSPTPFDRNFA	---TKGAKAMNMMSCKIKESYRNGRIFANTPDSGC	7088
574	QY	GLGVR	---WFSPLAPSFEDIAGYCHDKKIRWHISL	605
709	Db	VLCMRKRALYFQPVTELKQOTDFEHRIPKEQMWLKL	744	
RESULT 15				
FPVA_PSEAF				
ID	FPVA_PSEAE	STANDARD:	PRT;	815 AA.
AC	P48632; Q51339;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Ferripyoverdine receptor precursor.			
GN	FPVA OR PA3398.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
ON	NCBI_TaxID=287;			
OX	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.			
RC	STRAIN=CD10;			
RC	MEDLINE=93328663; PubMed=8335619;			
RA	Pooler K., Neshat S., Krebs K., Heinrichs D.E.;			
RT	"Cloning and nucleotide sequence analysis of the ferripyoverdine			
RT	receptor gene fpva of Pseudomonas aeruginosa.";			
RL	J. Bacteriol. 175:4597-4604(1993).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PAO;			
RL	Lamont I.L.;			
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 15692 / PAO1;			
RC	MEDLINE=20437337; PubMed=10984043;			
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goutry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Saller M.H., Hancock R.E.W., Lory S., Olson M.V.;			

RT *Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.*;
RL Nature 406:959-964(2000).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
CC STARVATION CONDITIONS.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L10210; AAA25819.1; -;
DR EMBL: U07359; AAB60199.1; -;
DR EMBL: AF004666; AAG05786.1; -;
DR PIR: A40601; A40601.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
FT SITE 798 815 TONB C-TERMINAL BOX.
FT CONFLICT 716 716 Y -> F (IN REF. 1).
FT CONFLICT 745 746 MISSING (IN REF. 1).
FT CONFLICT 750 750 S -> R (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A1IAE4AA5A290F35 CRC64;

Query Match 3.7%; Score 118.5; DB 1; Length 815;
Best Local Similarity 19.5%; Pred. No. 3.8;
Matches 112; Conservative 68; Mismatches 192; Indels 201; Gaps 30;
QY 66 DMVEEHLPLITQOEEVLKEQTGFLEAEP-----DNVKTMLRSGKGFSSKV--SLFE 117
DB 186 DVMRHRTPGIT--VSAYDTRNNYARGFSINNFOYDGPSTARVNGYSAGNTLSDMAI 242
QY 118 KCGAYTVHITPGPRTKIANVGVAIGLDLSDGNLAEEYRNALENWQOPVGS-----DFDQ 172
DB 243 YORVEVLKATGLTGAGSLGATI-----NLIR-----KKPHEFKGHVELGA 285
QY 173 DSWEN--SKTSVLGAVTRKGYPLAKLGNTRAA-----PLTESGNYRGRAVAAYQDKHSPMDHYERTKTSVYVYGTLEF 338
DB 286 GSWDNYRSELDVSG-----PLTESGNYRGRAVAAYQDKHSPMDHYERTKTSVYVYGTLEF 338
QY 203 -VNPDTATVDLNVVD-----SCRPIAGOFETGTQRYPEQIVSGLARQPGTPYD 253
DB 339 DLNPDTM--LTVGADYQDNDPRGSGWSGFLPDSQGNRNDVSRFSNNGAKWSSWEQYT 395
QY 254 LDLLDFQOALEQ-----NGHYSGASVQADFDFLQCD-----RVPKVSV 293
DB 396 RTVFANLEHFANGWGVKGVOLDHKNHGYH-----APLGAIMGDWPAPDSNAKIVAQKYT 449
QY 294 TEVKRKLE---TG---IRLSEYGLGKGIAYDYVNLNFKNGYIGSVVMDMDKYETTLAA 346
DB 450 GETKSNLDIYLPGPQFLGREHLYVGTSAFSHWE-----GKSYWNLNRYDNTTDD 502
QY 347 GIS-----QPNRY-----RGNVWTSNVSNRSTTONLEKRAFSGGIWYVDR 388
DB 503 FINWDGIDGRPDWGTSPQYIDDKTRQLGSYNTAR--FNVTDLNLN-----FLGG----- 549
QY 389 AGTDAR---LGAEFLAERKIP--GSDIDLGNHAT-----MLTASHKROLLNNVLH 435
DB 550 RVVDYRVTLNPTITRESGEIPYVGVAYDLDNTYSVYASVTDIFMPQDQSWYRDSNNKLE 609
QY 436 PE-----NGHYLDGKIGTTLGLTFLSSTALIRTSARAGYFTTPENKKLGTFLIRG-- 484

Db 610 PDEGQNYEIGKIGKVELDGLRNTSLAVF-EIHEENRAEEDALNSKPTNPAL-TYAYKGIK 667
QY 485 -----QAGYT--VARDNA 495
DB 668 AKTKGYEAEISGELAPGWQVQAGYTHKIIRDDS 700
RESULT 16
120K_RICRI STANDARD: PRT: 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia;
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=25154118;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT 120 kDa surface-exposed protein of Rickettsia rickettsii.;
RL Mol. Microbiol. 3:1579-1586(1989).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL: X16353; CAA34402.1; -;
DR PIR: S07575; S07575.
DR InterPro: IPR003858; rOmpA_rOmpB.
DR Pfam: PF02708; rOmpA_rOmpB; 1.
KW Antigen; Glycoprotein; Cell wall; s-layer.
FT CARBOHYD 7 7 N-LINKED (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).

```
FT CARBOHYD 861 879 N-LINKED (POTENTIAL).
FT CARBOHYD 879 899 N-LINKED (POTENTIAL).
FT CARBOHYD 920 926 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1211 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match 3.7%; Score 116.5; DB 1; Length 1300;
Best Local Similarity 19.68; Pred. No. 9.7;
Matches 117; Conservative 92; Mismatches 248; Indels 139; Gaps 24;

Qy 72 LPLITQOQEEVLD-----KEQTGLAEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVH 125
Db 219 IKLTSTQNNIVVDFDLATDQTGVVDASSLTNAQTL-----TINGKIG----- 262

Qy 126 ITPGPRTK-IANVGVAITLGDILSDGNLAAYRYNALENWQOPVGS-----FDQDSWENSKT 180
Db 263 -TIGANNKTLGQFNIGSSKTVLSNGNVA---INEL-----VIGNDGAVQFAHDTYLTIRT 313

Qy 181 S-----VLGAVTRKGYPLAKLGNTRAANVP-----DTATVDLANVVDSDGRPIAFGD 226
Db 314 TNAAGQCKIIFNPVNNGTLLAAGTNGSATNPLAEINFGSKGVNVDVTLNVGEGVNLIA 373

Qy 227 FEITGTQRY-----PEQIVSGLARFQGTPTDLDLL-----L 258
Db 374 TITTTDANVGSFVFNAGGTNIVSGTVGGQGNKFNVALENGTTVKFLGNATFNGNTTI 433

Qy 259 DFOQALQNGHYSGASV-QADFRLQGRDVPVKVSVTEVRKHLETKGIRLDSEYGLGCKI 317
Db 434 AANSTLQIGNNTADCVASADGTGIVEFVNTGPTITVTLNKQAAFPVNAKQITVSGPGNV 493

Qy 318 AYDYN-----LFNKGYIGSVVMDMDKYETTLAAGISQPNRYGNWYTSNVSY 365
Db 494 INEIGNAGNHGAVTDTIAENSLSGAVP-----LPRGI--PFNDAGNTWPLTI-- 541

Qy 366 NRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLAEGRKIPGSD-----IDLGNSHATML 421
Db 542 -KSTVGNKTAQKFDVPSVVV---LVGDS-----VIADQVIGDQNNIVGLGSDNGIIV 592

Qy 422 TASWKROLLNNVLPENGHYLDGKIGTTLGTSSTALIRTSARAGYFFTPENKKGTFI 481
Db 593 NATLYAGISTLNNQGTVTLSGVPNTPGTVYGLGTGIGASKFKQVTFTDYNLGNII 652

Qy 482 IRGQAGTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNGS-----VLPERA 533
Db 653 -----ATNATINDGVTVTTGGIAGI-GPD-GKITLGSVNGNVRFAFGDILSNST 700

Qy 534 LLVGSLEYQLPFTTLTSGAVFHDMDA---AANPKRMKLRKHGSLGVRWFSPLAPE 586
Db 701 SMIGTTKANNGTVTYLGNAFVGNIGSDSTPVSARETGSDSGAGLQGNIVYSOVIDF 756

RESULT 17
OMPB_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptidel.
GN OMPB
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RN SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Josse N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -I- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -I- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16353; CAA34403.1; -
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDa BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 3.7%; Score 116.5; DB 1; Length 1654;
Best Local Similarity 19.68; Pred. No. 14;
Matches 117; Conservative 92; Mismatches 248; Indels 139; Gaps 24;

Qy 72 LPLITQOQEEVLD-----KEQTGLAEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVH 125
Db 573 IKLTSTQNNIVVDFDLATDQTGVVDASSLTNAQTL-----TINGKIG----- 616

Qy 126 ITPGPRTK-TANVGVAITLGDILSDGNLAAYRYNALENWQOPVGS-----FDQDSWENSKT 180
Db 617 -TIGANNKTLGQFNIGSSKTVLSNGNVA---INEL-----VIGNDGAVQFAHDTYLTIRT 667

Qy 181 S-----VLGAVTRKGYPLAKLGNTRAANVP-----DTATVDLANVVDSDGRPIAFGD 226
Db 668 TNAAGQCKIIFNPVNNGTLLAAGTNGSATNPLAEINFGSKGVNVDVTLNVGEGVNLIA 727

Qy 227 FEITGTQRY-----PEQIVSGLARFQGTPTDLDLL-----L 258
Db 728 TITTTDANVGSFVFNAGGTNIVSGTVGGQGNKFNVALENGTTVKFLGNATFNGNTTI 787

Qy 259 DFOQALQNGHYSGASV-QADFRLQGRDVPVKVSVTEVRKHLETKGIRLDSEYGLGCKI 317
Db 788 AANSTLQIGNNTADCVASADGTGIVEFVNTGPTITVTLNKQAAFPVNAKQITVSGPGNV 847

Qy 318 AYDYN-----LFNKGYIGSVVMDMDKYETTLAAGISQPNRYGNWYTSNVSY 365
Db 848 INEIGNAGNHGAVTDTIAENSLSGAVP-----LPRGI--PFNDAGNTWPLTI-- 895

Qy 366 NRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLAEGRKIPGSD-----IDLGNSHATML 421
Db 896 -KSTVGNKTAQKFDVPSVVV---LVGDS-----VIADQVIGDQNNIVGLGSDNGIIV 946

Qy 422 TASWKROLLNNVLPENGHYLDGKIGTTLGTSSTALIRTSARAGYFFTPENKKGTFI 481
```


Db 947 NATLYAGISTLNNOGVTLSGGVPTPGTYGLGTGASFKQVTFDTTNNLGNII 1006
 QY 482 IRGAGCTVARDNADVPGLMFRSGASSVRGYELDSIGLAGPNS-----VLPERA 533
 Db 1007 -----ATNATNDGVTVTTGGIAGI-GFD-KKITLGSVNGNVRFPADGILNST 1054
 QY 534 LLVGSLEYQLPFTRTLSGAVFHDMDGA---AANFKRMKLKHGSLGVRWFSPILAPF 586
 Db 1055 SMIGTfKANNGTGYTLGNAPVGNIGSDTPVASRVFTGSDSGAGLOGNIYSQIDF 1110

RESULT 18

PMPL_CHLPN
 ID PMPL_CHLPN STANDARD; PRT; 922 AA.
 AC Q929C5: Q9K1Y9; Q924H9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Probable outer membrane protein pmpl precursor (Polymorphic membrane
 DE protein 1) (Outer membrane protein 6).
 GN PMPL1 OR OMP6 OR CPN0005 OR CP0770.
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VR1310.
 RX MEDLINE=20007584; PubMed=10539856;
 RA Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygind P.,
 RA Madsen A.S., Knudsen K., Falk E., Birkelund S.;
 RT "Molecular biology of Chlamydia pneumoniae surface proteins and their
 RT role in immunopathogenicity".
 RL An. Heart J. 136:S491-S495(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis".
 RT Nat. Genet. 21:385-389(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39".
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA".
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC
 DR EMBL: AJ133035; CAB37075.1; -
 DR EMBL: AE001585; AD18163.1; -
 DR EMBL: AE002237; AAF38570.1; -
 DR EMBL: AP002545; BAA98215.1; -
 DR TIGR: CP0770; -
 DR InterPro: IPR003368; Chlamydia_PMP.
 DR InterPro: IPR003357; OMP.
 DR Pfam: PF02385; OMP; 1.
 DR Pfam: PF02415; DUF145; 1.
 KW Outer membrane; Signal; Multigene family; Complete proteome.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 922 PROBABLE OUTER MEMBRANE PROTEIN PMPL.
 FT CONFLICT 14 14 F -> L (IN REF. 1).
 FT CONFLICT 375 375 Y -> C (IN REF. 1).
 FT CONFLICT 606 606 D -> N (IN REF. 1).
 SQ SEQUENCE 922 AA; 100457 MW; DFF2AB633AB031C CRC64;

Query Match 3.7%; Score 116; DB 1; Length 922;
 Best Local Similarity 18.5%; Pred. No. 6.5;
 Matches 131; Conservative 83; Mismatches 226; Indels 268; Gaps 29;

QY 108 YFSKVSLETKDGAYTVHTTPGPRTKIANVGVAILDGLSDGNLAE----YYRNAL---- 159
 Db 94 YFNMISSGTTKEGA--VLCCQDPQATARFSGFTLSFIQSPGDIKEQGCLYSKNALMLLN 151
 QY 160 -----ENWQOPVG-----SDFDQDSWENSKTSVLGAVTRKGYPL----- 193
 Db 152 NYVVRFEQNOQSKTKGGAIGANVTIVGNYSVSFYQNAATFGGAIHSSG-PLQIAVNOAE 210
 QY 194 -----AKLGNTRAAVNPDTATVDLVNVVDSGRPTAFGDFEITGTQRYPEQIVSGLARF 246
 Db 211 IRFAQNTAKNGSGALYS-----DGDIDIDQNAVYLVFRENEALTATGAGGAVCCCLPTS 264
 QY 247 QPCTPYDLDLLDFQQALEQNGH--YSGASVQADFRLQDGRVVPVSVTVKRRHKLETG 304
 Db 265 GSSTPPIVTFSDNKQLVPERNHSIMGGGAIYAR-----KLSSSGGPTLPINN 313
 QY 305 IRLDSEYGLGGKAYD-----IRLDSEYGLGGKAYD----- 320
 Db 314 ISYANSONLGGATAIDTGTGGEISLSAEKGTITFGQNRTPSLPLNGIHLLQNAKFLKQARN 373
 QY 321 -----YYNLF-----NKGYSVVMMDKYTTLAAGISQPRNRYGNY 358
 Db 374 GYSIEFYDPTTSADGSTQNLINGDPKNKEYTGTILESGEK---SLA---NDPRDFKS-- 425
 QY 359 WTSNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDAR-----LCAEFLA--EG 403
 Db 426 -----TIPQNVN--LSAGYLVVIKEGAETVSKFTQSPGSHLVLDLGLTKLIASKED 473
 QY 404 RKIPGSDIDLGNSHATMLTASWKRQLLN-----NVLHPENGHYLD----- 443
 Db 474 IAITGLAIDIDLSLSSSTAIVIRANTANKQISVTDSITELISPTGNAYEDLMRNSQTFPL 533
 QY 444 -----GKIGTTLGTFLS-----STALITTSARAGYFF-----TPENKKLG 478
 Db 534 LSLEPGAGGSVTVTAGDFLPVSPHYGQGNWKLAWTGTGNKVGFEFFWDKINYPKPEKEG 593
 QY 479 TF---IIRGOAGTYVARDNADVPGLMFRSGGASSV---RGYELDSIG----- 520
 Db 594 NLVPNILMGNA-----VDVRSMLQVQETHASSLQTDRLGLWIDGIGNFHVFSASEDNI 645
 QY 521 -----LAGPNGSVLP-----ERALLVGSLEYQLPFTRTLSGAVFHDMD 557
 Db 646 RYRHNSGGYVLSVNSNITPKHYTSMAPSFOLSRDQKAVSNNEY-----RMYLGSYLYQY 700
 QY 558 GDAANFKRMKLKHGS---GLGVRWF--SPLAPSFDAIAYGHSDKKIR 600
 Db 701 TTSLGNIFRYASRNPNVNVGILSRFLQNPMLPFHFLCAYGHATNDMK 748

RESULT 19

APLP MANSE
ID APLP MANSE STANDARD; PRT: 3305 AA.
AC Q25490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Apolipophorin precursor [contains: Apolipophorin-I (APOLP-1);
DE Apolipophorin-II (APOLP-2)].
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Spingioidea; Spingidae; Spinginae; Manduca.
OX NCBI_TaxID=71130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fat body;
RX MEDLINE=97166603; PubMed=9014323;
RA Sundermeyer K., Hendricks J.K., Prasad S.V., Wells M.A.;
RT "The precursor protein of the structural apolipophorins of
RT lipophorin: cDNA and deduced amino acid sequence.";
RL Insect Biochem. Mol. Biol. 26:735-738(1996).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; SECRETED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U57651; AAB53254.1; -
DR InterPro: IPR001847; LipID_transprt_N.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF00094; vwd; 1.
DR Pfam: PF01347; Vitellogenin_N; 1.
DR SMART: SM00216; VWD; 1.
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 ? APOLIPOPHORIN-II.
FT CHAIN ? 3305 APOLIPOPHORIN-I.
SQ SEQUENCE 3305 AA; 366941 MW; B27D611410285FD7 CRC64;

Query Match 3.68; Score 114.5; DB 1; Length 3305;
Best Local Similarity 19.48; Pred. No. 48;
Matches 118; Conservative 106; Mismatches 232; Indels 153; Gaps 32;

Qy 28 ENKAAGFALFRN---KSPDTSVKLKPFPVPRIDTQDSEIKDMVEEHLPLITQOOEEVLD 84
Dy 1286 QDKSAGFNLDARTHKREGDYSIKVKNL-----NANLEAFSRDIIVN 1329
Qy 85 KEQTGFLEAEAPDNVKTMLRSKGYFSSKVSUTEKGAYTVHTPGPTKIANVGVAIGD 144
Dy 1330 AEKS-----NVENIDMKGVGRVLS-----GFVLH-----NTPKPDVNVGVFIH 1369
Qy 145 I-LSDGNLAERYRANLENWQPVGSDPDQDSWENSKTSVLGA-----VTRKGYPL 193
Dy 1370 LKINGKNEFDKINIGHIETP-----AVFSSHATISGRGDIIDYLLKIMRTAMP- 1420
Qy 194 AKLGNTRAAVNPDTATVDLNVVDS---GRPIAFGDFEITGTQRYPEQIVSGLARFPGT 250
Dy 1421 --NGNEFLVIKDSIAANGQYKVTADGKNGLLIIDEFKINRK-----IKGDVRFYTAKE 1472
Qy 251 PY---DLDLLLDFOALEQNGHYSGASVQADFRLQGRVPVKVSVTEVRHK-----LET 303
Dy 1473 PVFNADIDLFLNFKEKNSDKVHFSTYNNKTKDKVMDTKNLEYACKRTEVNIHQDGILAVT 1532
Qy 304 G-----IRLDSYGLGGKTAID-----YYNLFNGYIGSVVWMDKVTETTLAAGISQ 350
Dy 1533 GKAITVAELVLPTRCRLSLKIDHGAQFKDGLYN-----GHMDMTISDAPK----- 1577
Qy 351 PRNYRGNYWTSNVSY--NRSTQTNLEKRA--FSGGI--WVVRDRAGIDARLGAFLAEGRKI 406
Dy 1577

Db 1578 ----RGS-GASTISYKGVSNLNLDQEIIDYEGQINFKLKDKGNLQSTFSLKNNPDGDKF 1632
Qy 407 P----GSDIDLGN---SHATMLTASWKROLLNNVLPENCHYLDGKIGTGTGLTFLSSTALI 460
Dy 1633 KYEFKSDVN-GNLIPKANLVATGYYSNSEI---DETYRLKGSYSDIGFELAGVTI 1688
Qy 461 RTSARAGYFFTPENKKLGTFIIRGOAGYTVARNADVPGLMFRSGG-----ASSVR 512
Dy 1689 K-----FLDAGDKKYLDDYTLTVRLPFEKAHDIKWSTVLFLOPQGOEMTEYTLVESVQ 1742
Qy 513 ---CYELDSITGLACP---GSVLPERALLVGSLEYQLPFTRTLSGAVPHDMGDAANPK 565
Dy 1743 INADYIKIDANGKVGPKNGYGAV----KVLVPHVE---PFVLDYNYKSSHE-GEKNNYV 1794
Qy 566 RMKLKHGSG 574
Dy 1795 ELKTKYKGK 1803

RESULT 20
VGNM_BPWV
ID VGNM_BPWV STANDARD; PRT: 1018 AA.
AC P23009;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein M [Contains: Coat protein VP37; Coat protein VP23].
OS Bean-pod mottle virus (strain Kentucky G7) (BPWV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Comovirus.
OX NCBI_TaxID=31715;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91272504; PubMed=2053290;
RA Macfarlane S.A., Shanks M., Davies J.W., Zlotnick A.,
RA Lomonosoff G.P.;
RT "Analysis of the nucleotide sequence of bean pod mottle virus middle
RT component RNA.";
RL Virology 183:405-409(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 447-1018.
RX MEDLINE=89317459; PubMed=2749253;
RA Chen Z., Stauffer C., Li Y., Schmidt T., Bomu W., Kamer G.,
RA Shanks M., Lomonosoff G., Johnson J.E.;
RT "Protein-RNA interactions in an icosahedral virus at 3.0-A
RT resolution.";
RL Science 245:154-159(1989).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF THE TWO COAT
CC PROTEINS.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-103 IS THE
CC INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M62736; AAA42801.1; -
DR PIR: A40321; GNWGX7.
DR PDB: 1BMV; 15-OCT-91.
DR InterPro: IPR003181; Como_LCP.
DR InterPro: IPR003182; Como_SCP.
DR Pfam: PF02247; Como_LCP; 1.
DR Pfam: PF02248; Como_SCP; 1.
KW Coat protein; Glycoprotein; polyprotein; 3D-structure.
FT CHAIN 447 820 COAT PROTEIN VP37.
FT CHAIN 821 1018 COAT PROTEIN VP23.
FT DOMAIN 447 628 C DOMAIN.


```

-!- FUNCTION: COULD BE INVOLVED IN THE EXPORT AND ASSEMBLY OF THE
CC CC PUTATIVE VRAH FIMBRIAL SUBUNIT ACROSS THE OUTER MEMBRANE.
CC CC
-!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC CC (by similarity).
CC CC
-!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC CC
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
-----
EMBL; U18997; AAA57947.1; ALT_INIT.
DR DR
EMBL; AE000395; AAC76178.1; -.
DR DR
Ecogene; EG12775; Vraj.
DR DR
InterPro; IPR000015; Fimb_usher.
DR DR
Pfam; PF00577; Usher; 1.
DR DR
PROSITE; PS01151; FIMBRIAL_USHER; 1.
DR DR
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbriae;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 838 YRAJ.
FT DISULFID 815 837 POTENTIAL.
FT SEQUENCE 838 AA; 93616 MW; FD265CC903A269BA CRC64;
SQ
Query Match 3.6%; Score 113; DB 1; Length 838;
Best Local Similarity 18.3%; Pred. No. 8.8;
Matches 106; Conservative 63; Mismatches 166; Indels 244; Gaps 28;

QY 141 ILGDILSDGNLAEYR-----NALENWQPVGSDFDQDSWENKTSVLGAVTRKG 190
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 254 LGESATSGSISSFTTFCVQLASDNNLPSQR---GFAPTVRGIRANSAIVTRQG 309
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 191 YPLAKLGNTRAAVNPDTATVDLNVVDSGRPTAFGDFEIT-----GTQRYPEQIVSGLAR 245
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 310 YVIYQ-----SNVSAGAFEINDLYPPSNS--GDLEVTIEESDGTOR-----R 349
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```
QY 246 F-QGTPYDLDLLDFOQALEQNGH---YSGASVQADFDRLQGRVVPVKVSVTEV-----KR 298
Db 350 FIQPYSSLPN-----MQRPHLKYSATAGRYRADANSDSKEPEFAEATAIYGLNNT 400
QY 299 HKLETGIRLDEY-----GLGG----- 315
Db 401 FTLYGGLGSEDYALGIGIGTGLGALGALSMDINRADTQFDNQHSFGHWRTQYIKDI 460
QY 316 -----KIADYYNLFNKGYYI-----GSVVWMDKYE-----TTLAAGIS-----QP 351
Db 461 PETNTNIAYSVYRYTNDGYSEFNEANTRNDYNSRQKSETOFNLSQIFDGVSLYASGSQ 520
QY 352 RNRGNV-----WTSNVSYNRSTQNLKRAFSGGI-----WYVRD 387
Db 521 QDYMGNNDKNNRISVSGVGQWGVGYSLNQYRYTQDN-NDRAISLNLSTPLERWLP 579
QY 388 RAGI-----DARLGAELFAGRKIPG--SDIDLGNSHATMLTASWKRQL----- 429
Db 580 RVSYQMTSQDRPTQHEMRLOGSLDDGRLSYSLEQSLDDNNHNSLSNAYSRYSPYGTFS 639
QY 430 ----LNN-----VLHPENGHYLDGKIGTTLGTFLSST-ALIRTSARAG----- 467
Db 640 AGYSYGNDSQYNGVGTGVVIPH-----GVTLSQYLGNAFALIDANGASGVRIQ 690
QY 468 -----YFFTPENKKLGT-----FIIRGQAGYTVARDNA 495
Db 691 NYPGIATDPFGYAVVPYLTQYQENRLSVDTQLPDNVDLEQTQFVVPNRMGAMVAARFNA 750
QY 496 DVPSGLMFRSGGASSVRYGELDSIGLAGPNGSVLPERAL 534
Db 751 NI-----GYRV-LVTVSDRNGKPLPFQAL 773

RESULT 23
VP1_BCHP
ID VP1_BCHP STANDARD; PRT; 595 AA.
AC F19192;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Protein VP1 (ORF1).
OS Bacteriophage Chpl.
OC Viruses; ssDNA viruses; Microviridae; Microvirus.
OX NCBI_TaxID=12367;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RX MEDLINE=90111716; PubMed=2607341;
RA Storey C.C., Lusher M., Richmond S.J.;
RT "Analysis of the complete nucleotide sequence of Chpl, a phage which
RL Infects avian Chlamydia psittaci."
RL J. Gen. Virol. 70:3381-3390(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00624; BAA00515.1; -
CC DR PIR; J00345; J00345.
CC KW Structural protein.
CC FT INIT_MET 0 0
CC FT SIMILAR 16 73 TO F PROTEIN OF PHAGE PHI-X174, AA
CC FT SEQUENCE 595 AA; 66850 MW; AE71D8024EA967ED CRC64;
CC RESIDUES 2-58.

Query Match 3.5%; Score 111.5; DB 1; Length 595;
Best Local Similarity 23.5%; Pred. No. 6.8; Indels 79; Gaps 22;
Matches 107; Conservative 55; Mismatches 214;

QY 14 FFFPHAYAPAADLSENKAAGFALFNKNSPDTESVKLKPFPFRIDTQDSIKDMVEEHL 73
```

```
Db 117 YFVPQVTSPOGGYAENSIYDFGLPTKVANYRH-QVLPRLAYNLIFENYRDENLQESLP 175
QY 74 LITOOOEVLDKQETGFLAEE---APDNVKTMLRSK--GYFSSKVSILTERDQATVHTTP 128
Db 176 VMTGDADPKVD-PTTGEESEDDAVPYVYKLMRRNKRYDFTSALPLGLOKGPSVGIGITG 234
QY 129 GPRTKIANVGVAI---LGDILSDG-NLAEEYRNALNNQOPVGSDFDQDSWENSKTSVLG 184
Db 235 GDSCLPLVHGLAIRSYLDDSSDDQSFSGVSIVNASQKW-----FTAD---GRITSGMG 284
QY 185 AV---TRKGYP-----AKLGNTRAAVNPDTATVDLNVVVDVSGRPTAFGDFEITGTRY 235
Db 285 SVPGTGTGTFPIDNVVPSYFGTTVAQTGSPSS-----STPPFVKGDFP-----Y 331
QY 236 PEQIVSGLARFQCTPYDLDLDDFOQALEQNGHSG---ASVQADFDRLQD---RVVPV 289
Db 332 VD-----LAASSVTINSLRNATLQOMFEKSAARYGSRYVESVQGHFVHGLDYRAORPI 386
QY 290 -----KVSVEYKRHKLETGIRLDSFYGLGGKIAY-----DYNLNFNKGYYI---GSVVWD 336
Db 387 YLGSKSYSVSNVNVQNS-----STDSVSPQGNLSAYALSTDTKHLFTKSFVHGFIYGL 441
QY 337 MD-----KYETTLAAGISQPRNYRGNWYTSNVSYNRSTQNLKRAFSGGIWYVRDRAGI 391
Db 442 LSATADLTQOGLERQWSRFSRY-DYWPMTFAHLGEQPVYNKEIYCQSDTVMDFSPGSASN 500
QY 392 DARLG-AEFLAEGRKIPGSDIDLGNSHATMLTASW 425
Db 501 DVPPGYQERYAEYRYKPSKVTGLFRSNATGTLDSW 535

RESULT 24
TACY_STRCB
ID TACY_STRCB STANDARD; PRT; 574 AA.
AC Q53957;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Streptolysin O precursor (Thiol-activated cytolysin).
GN SLO.
OS Streptococcus canis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95102113; PubMed=7803818;
RA Okumura K., Hara A., Tanaka T., Nichiguchi I., Minamide W.,
RA Igarashi H., Yutsudo T.;
RT "Cloning and sequencing the streptolysin O genes of group C and group
RT G streptococci".
RL DNA Seq. 4:325-328(1994).
CC -!- FUNCTION: SULFHYDRYL-ACTIVATED TOXIN. IS ABLE TO LYSE CHOLESTEROL
CC CONTAINING MEMBRANES. CAN BE REVERSIBLY INACTIVATED BY OXIDATION.
CC CHOLESTEROL IS THE RECEPTOR FOR THE BINDING OF THESE TOXINS TO
CC EUKARYOTIC CELL MEMBRANES.
CC -!- SIMILARITY: BELONGS TO THE THIOL-ACTIVATED CYTOLYSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D16825; BAA04105.1; -
CC DR HSSP; P19995; IPFO.
CC DR InterPro; IPR001869; Thiol_cytolysin.
CC DR Pfam; PF01289; Thiol_cytolysin; 1.
CC DR PRINTS; PR01400; TACYTOLYSIN.
CC DR ProDom; PD007062; Thiol_cytolysin; 1.
```


QY 316 KIAYDYNNLFPN-KG-YIGSVWMDMKYETTLAAGISOPRNYRGNWTSNVSYNRSTQNL 373
 DB 366 RVANGQVQIFSGRGTAGT-----DSSLTNIRAPLQFSGPALTASLGSGPITYNS 417
 QY 374 EKRAF-----SGGIWYVRDRAGIDARLCAEF-----LAERKIPGSDIDLG-----NS 416
 DB 418 NNGTFLSGTSGPMWVDONRQVNPAGLVFOGNNLVPNLADPLAISDSKISLSLGPGLTQ 477
 QY 417 HATMLTASMKROLLNVLHPENHYLDGKIGTTLGTFLSTALIRTSARAGYFFETPENKK 476
 DB 478 ASNALTSL-----LGNLEFSN-QAVAIKAGRCL-RFESSQALSSLVNGNLT----- 525
 QY 477 LGTFIIRGAGYTV-ARDN---ADVPGLMFRSGGASSVGYELDSIGLAGPNSGLP-- 530
 DB 526 LDTVIRPNLGDGLEVRDNKIIKVKANLRFENGAVTA-----GTVNPSAPEAPT 576
 QY 531 ---ERALLVGSLEYQLPFTTILSGAVFHD-----MGDA-AANFKRMKLKHGSLGVR 578
 DB 577 LTAEPPLRASNSHLQSLSL-----EGLVYHNNALALQLDGMEVNHQGLTLRVGSGLQMR 631

RESULT 26
 YH3G_ECOLI
 ID YH3G_ECOLI STANDARD; PRT; 691 AA.
 AC P37645; P37644;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yhjg.
 GN YHJG OR B3524.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes.;
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [2]
 RP REVISIONS.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U00039; AAC18501.1; ALT_FRAME.
 DR EMBL: A5000429; AAC76549.1; -
 DR EcoGene: EGI2251; yhjg.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 652 672 POTENTIAL.
 SQ SEQUENCE 691 AA; 75130 MW; CFACD981B135F4C3 CRC64;

Query Match 3.5%; Score 110.5; DB 1; Length 691;
 Best Local Similarity 20.0%; Pred. No. 9.8;
 Matches 96; Conservative 53; Mismatches 143; Indels 187; Gaps 23;

QY 161 NMQOPYGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGT-----RAAVNPDTATVDLNVV 215
 DB 144 NWTFLNANDNDKDA--NAKPSAMS-----FRLDNILFDQGRITADDKVSKADLEIF 192
 QY 216 VDS-GRPIAFGDEITGTQRYPEQIVSGLARFOPGPPYDLDDLLDFQOALEQNGHYSAS 274
 DB 193 VDPGLKPLPFS--EVTGSK-----GKA-----DKEKGVGYFGLKAGQRYNG-- 232
 QY 275 VOADFRLQGRDVRPVKVSVEVKRHKLETGIRLDSEYGLGGKIAYDYNNLNFNGYIGSVV 334
 DB 233 -----EPLTG-----TG-----KIGMLA----- 246
 QY 335 WMDKYETTLAAGISOPRNYRGNWTSNVSYNRSTQNLKRAFSGGIWYVRDRAGIDAR 394
 DB 247 -----LRGEGTFFPVQADFRCNT-----RVAFDGVVNDPMKMGGVDLR 285
 QY 395 LGAEFLAERKIPGSDIDLGNSHATMLTASWKROLLNNVLHPENHY-LDGKIGTTLGTG 453
 DB 286 L-----KFGSD--SLGDLYE-----LTGVLLPDTPTTPTDGR----- 316
 QY 454 LSSTALIRTSARAGYFFETPENKKLTGTFIIRGQAGYTVARD-----NADVPS----- 499
 DB 317 ---VAKIDTEKSSVFDYRGFNGRIGSDINGSLVYTTGKPRKLEGVESRQLRLADLGP 373
 QY 500 --GLMFRSGGASSVGYELDSIGLAGPNSGLVPERALLVGSLEYQLPFTTILSGAVPHDM 557
 DB 374 LIGVDSCKGAEKSKRSQKKGKSVQPAKV-----LPYDRFETDK--WDV 417
 QY 558 GDAANFKRMKLKHGSLGVRWFSF-----LAPFSFDIAYG-----HSDKK 598
 DB 418 MDADVRFKGRRIEHSGLSPIISLSTHILKNADLRQLQKFGMAGGSAIAIHLEGDKK 476

RESULT 27
 MRKC_KLEPN
 ID MRKC_KLEPN STANDARD; PRT; 828 AA.
 AC P21647;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane usher protein mrkC precursor.
 GN MRKC.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IA565;
 RX MEDLINE=91100388; PubMed=1670938;
 RA Allen B.L., Gerlach G.-F., Clegg S.;
 RT "Nucleotide sequence and functions of mrk determinants necessary for
 RT expression of type 3 fimbriae in Klebsiella pneumoniae.";
 RL J. Bacteriol. 173:916-920(1991).
 CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE TYPE 3
 CC FIMBRIAL SUBUNIT (MRKA).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 CC (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M55912; AAA25095.1; -
 DR PIR: D39142; D39142.
 DR InterPro: IPR000015; Fimb_usher.
 DR Pfam: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.

```
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 828 OUTER MEMBRANE USHER PROTEIN MRK.
FT DISULFID 813 827 POTENTIAL.
SQ SEQUENCE 828 AA; 91049 MW; B30EDF5798249FC9 CRC64;

Query Match 3.58; Score 110; DB 1; Length 828;
Best Local Similarity 19.8%; Pred. No. 14;
Matches 127; Conservative 84; Mismatches 213; Indels 216; Gaps 36;

QY 55 VRIDQSEIKDNVEEHLPLITQOEEVLDKEQTGFLLA-----EAPDNVKTM- 102
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 VNLATDDNMFPDGRSAP-----EIRDVAQSNALATVRQGSNIIVQTTPGPTLQ 319
QY 103 -LRSGYFSS-KVSLTERDKGAYTVHITPGPRTKIANVGVAITGLDLSGNLAEYRNA-- 158
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 DVPYSGYCDLEVSKEGDSVEFVSP-----YASVAQLLRG-WTRVALSACK 368
QY 159 -----LEN-----WQPPVGSDF-----DQDSWENSKTSVLGAVTRKGYPLAKLGNTRA 201
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 VDDSSLRNKPMLYQCTWQHGLNLLFTGYTGTGDDYQAPLLGTGMNTGICALSFDVTH 428
QY 202 AVNPDTATVDLVNVDSCRPTAFGDFEITGTQRYPE---QIVSLGARFPQPTDLDLLL 258
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 RLKSDT-----LDEHQ-----SYRATFNRMFTETQTSIVLAAYRYSTKGYNLN--- 473
QY 259 DFOQALEQNGHY-SCASVQADFRLQGRDVPVKYSVTEVRKHLETGIRLDSYGLGSKI 317
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 DALYAVDQEKYNSNYTWRO-----KNGMFTVNQNLPGDW---GGFYLCGRV 519
QY 318 AYDYNN-----LFNKGYIGSVVMDM-----KYETTLAAGISOP----- 351
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 A-DYNNRSGTEKQYOFYSNNMY-CRLSWSVDAQRYTPDSSGHRDRDRVSLNFSVPLWFG 577
QY 352 RNYRGNTWTSNVNRSSTONLEKRAFGGGIYWRDRAGIDARLGAE-----FLAE 402
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 ENRTANL-TSNTAFNNSR-----PASSQGVNGSLDSENNLNYGVSTTTAT 622
QY 403 GRKIPGSDIDLGNSHAYMLTASMKRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRT 462
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 GRQ---HDVALNGSYRPTWTL-----NGSYSGEGYRSGVGCASGTLI--- 663
QY 463 SARAGYFTPTPENKLGTFIIRQAGYTVARDNADVPGLMFRSGGASSV----- 511
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 AHQHGVSFSP-----TGPTMALTEAKDAAGVMLPGSPGTRIDNSGVAILPYL 711
QY 512 RGYELDSGLAGPNS-----VLPERALLVGSLEYQLPFTTLTSGAVFHDMD 559
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 RPYRINSVEI-DPKGSNDVAVGSTVAQVWPWEGSV-----KVSFDTTLQNNI----- 759
QY 560 AAANFKRMKLKHGSLGVRFSPSPLAPTSFOIAYGHSKKI 599
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 760 -----TLRARQANGL-----PL-PRAATI-FGPGSGKEI 785

RESULT 28
FRUA_STRMU STANDARD; PRT; 1423 AA.
AC 003174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-
DE fructosidase) (Fructanase).
GN FRUA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93014178; PubMed=1398976;
```



```
QY 409 SDIDLGSHATMTASWKKROLLNNVLPENHGLDGIKIGTTLGTFLSSTALIRTSARAGY 468
Db 1057 YIDL-----KYQRLIN-----LFVASGNIDPSQAYSQ 1086

QY 469 FFTPENKKLGTFIIRGOAGYTVARDN 494
Db 1087 FGDSETVRLYRF-----AGDTIAEAN 1107

RESULT 29
TACY_STRPY
ID TACY_STRPY STANDARD; PRT; 571 AA.
AC P2131;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Streptolysin O precursor (Thiol-activated cytolysin).
GN SLO OR SPY0167.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88057628; PubMed=3502717;
RA Kehoe M.A., Miller L., Walker J.A., Boulnois G.J.;
RT "Nucleotide sequence of the streptolysin O (SLO) gene: structural
RT homologies between SLO and other membrane-damaging, thiol-activated
RT toxins."
RL Infect. Immun. 55:3228-3232(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=SF370 / ATCC 700294 / Serotype M1;
RA MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvorov A.N., Kenton S., Lal H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- FUNCTION: SULPHYDRYL-ACTIVATED TOXIN. IS ABLE TO LYSE CHOLESTEROL
CC CONTAINING MEMBRANES. CAN BE REVERSIBLY INACTIVATED BY OXIDATION.
CC CHOLESTEROL IS THE RECEPTOR FOR THE BINDING OF THESE TOXINS TO
CC EUKARYOTIC CELL MEMBRANES.
CC -1- SIMILARITY: BELONGS TO THE THIOI-ACTIVATED CYTOLYSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18638; AAA26975.1; -.
DR EMBL; AE006485; AAK33267.1; -.
DR PIR; A43507; A43507.
DR HSP; P19995; 1pfo.
DR InterPro; IPR001869; Thiol_cytolysin.
DR Pfam; PF01289; Thiol_cytolysin; 1.
DR PRINTS; PR01400; TACYTOLYSIN.
DR PRODOM; PD007062; Thiol_cytolysin; 1.
DR PROSITE; PS00481; THIOI-CYTOLYSINS; 1.
KW Toxin; Hemolysis; Cytolysis; Signal; Complete proteome.
FT SIGNAL 1 33
FT CHAIN 34 571 STREPTOLYSIN O.
FT ACT_SITE 530 530 BINDING TO CHOLESTEROL (BY SIMILARITY).
SQ SEQUENCE 571 AA; 63638 MW; D05AA9979DCBACB0 CRC64;

Query Match 3.4%; Score 109; DB 1; Length 571;
Best Local Similarity 19.08; Pred. No. 9.4;
Matches 124; Conservative 92; Mismatches 234; Indels 202; Gaps 31;
```

```
QY 8 LLLPALFFPHAYAPAADLSNKAAAGFALFKNNKSPDTSVSKLKPFVRIDTODSEIKDM 67
Db 16 LLTAALLIIGNLVTAESNKONTASTETTTTNEQPKPESSELTTTEKAGQTDMLNSMD 75

QY 68 VE---EHLPLITQOOBEVL--DKRQTGFLAEEAPDNVKTMLRSKGYFSSKVSUTEKDGAY 122
Db 76 IKLAPKEMPLESAEEKSKEDKKKS---EDHTTEENDKIYSLNY--NELEVLAKNGET 130

QY 123 TVHITPGRTKIANVGVAIILGDILSDGNLAERYNALENMQQPVGSDGDDSWNSKNT-- 180
Db 131 IENFVPKEGVKKDKFIVI-----ERKKKNINTTP 160

QY 181 ---SVLGAVTRKGYPLA-KLGNTAAVNPDTATVDLVNVDVSGRPIAFGDFEITGTORYP 236
Db 161 VDISIIDSVDTRTPAALQANKGFETENKPDVV-----TKRNP 199

QY 237 EQI-----VSGLARFQPGTPYDLDLLDFOQALEQ-NGHYSGASQVADPRLOGDRVPV 289
Db 200 QKIHIDLPGMGDKRATVEVNDPTYANVSTADNLVNMHDNYSG-----GNTLPA 248

QY 290 KSVSTE---VKRHKLETGIRLDSEYGLGGKIAYDYNNLFNKG-----YIGSV 333
Db 249 RTQYTESWYYSKSOIEAALNVNSKI-LDGTIGIDFKSI-SKEKKVMIAAYKQIFVTVA 306

QY 334 -----VNDMD-KYETTLAAGISQPRNRYGNWTNVSYNRTSTONLEKRAFSGGIWY 384
Db 307 NLPNPADVEDKSVTFEKLQKGVY---NEAPPLFVSNVAYGRTVFVKLETSSKSNV-- 361

QY 385 VRDRAGIDARL-GAEFLAEGRKIPGSDIDLGN-----HATMLTASWKKROLL 430
Db 362 ---EAAFSAAKLGTDVKTNGKY---SDI-LENSSFYAVVLGGDAAEHNKVYTKDF--DVI 412

QY 431 NNVLHPENGVLDGKIGTTLGTFLSSTALIRTSARAGYFETPENKKGITFIIRGOAGYTV 490
Db 413 RNVI-----KDNATFSRKNPAYPIYTSV-----FLKNNKIAG---VNRTEY-V 453

QY 491 ARDNADYPSGLMFRSGGSSVRYGEL--DSIGLAGPNSVLPERALLVGSLEYQLPETRT 548
Db 454 ETTSTEVTSKGINLSHOGAYVAQVEILWDEIN----- 485

QY 549 LSGAVFIDMGDAANFRMKLHGSGLVGRWFSPLAFSPFIAYGHSDKKTR 600
Db 486 ---YDDKGKGVTKRRW-----DNNWYSKTSFSPFTVIPLGANSRNIR 524

RESULT 30
FHAC_BORPE
ID FHAC_BORPE STANDARD; PRT; 584 AA.
AC P35077;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hemolysin activator-like protein FHAC precursor.
GN FHAC.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA Willem R.J.L.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO P MIRABILIS HPMB AND S.MARCESCENS SHLB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64876; CAA46092.1; -.
DR
```

```
DR MEROPS: M19.UNW; -.
KW SIGNAL.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 584 AA; 64445 MW; 85625B2C07702178 CRC64;

Query Match
Best Local Similarity 3.4%; Score 109; DB 1; Length 584;
Matches 128; Conservative 71; Mismatches 238; Indels 186; Gaps 28;

Qy 22 PAADLS-ENKAAGPALFNKSPDTSVKLPKFPVRIDTODSEIKDMVEEHLPLITQOE 80
Db 64 PPVELNPOSEAAAPA-----RKPDATSGHTVTVHVAVDL---DFGVEGRFLDPAPLVQDYLN 116
Qy 81 EYLDKEQGTGFLAAEAPDNVKTMLRSKGYSFKSVLTER---DGAYTVHITGPRTKIANV 137
Db 117 RPLDNEQL-FLLVKA---LSAALYDRGYATSIVTFVPCVVDGVLLKLVENG-RIK----- 167
Qy 138 GVAILGDILSDGNLAEYR-----NALENQOOPVGSDFDQDSWENSKTSVLGAVTRKGY 191
Db 168 -----GWLIDGKPLEGTRDRMVFESAMPGWQDVLNVFDID----- 203
Qy 192 PLAKIGNTRAANPDATVNLNVVDSGRPIAFGDFEITCTORYPEQIVSGLARFQPGTP 251
Db 204 -----QAINYINNGKTNITIVPADEYGYSLDLQORRALPRVSLGMDNSGPGTP 255
Qy 252 YDLDLLLFOQALEQNGHYS-GASYQADFDRLQGDVRPVKVSVEVKRHKLETGIRLDSE 310
Db 256 -----ENGRYKYNASVTAN-----DLLGLNDTL-----GLYIGNR 285
Qy 311 YGL-GGKIAYDYNLFNKGYIGSVVWMDKYETTLAAGISQPRNY-----RGNWYT--SNV 363
Db 286 YYRDAGHDAERNYDLMYSVPLGTRLD-----LOTGYSTYRNLTKRYGQYOSAGNSR 338
Qy 364 SYNRTSTONLEKRAFSGGIWTVRRDAGIDARLGAERFLAEGRKIPGSDIDLGNSHATMLTA 423
Db 339 SFGLKATRLLYDRTRSQSVY-----GGUKLRQNKNYLA-----GTRLDVSKKHSYDVTV 388
Qy 424 SWKROLLNVLHPENGHYLD-----GKIGT-----TLGTFLLSITALIRTS 464
Db 389 GMYSTORGA---NAYFGDLSETFRGVGNVNGKYAAYDERGPGNVSRFENSLAWTRYMA 444
Qy 465 RAGYFFFTENKKGITFTIRGQAGTYVARDNADVPGLMFRSGGASSVRYEL-----DS 518
Db 445 LAGQ-----PIOWASQLGFQYSRQ--QLLSYQITVGVDEVTVRGYNLRTSQSGDS 492
Qy 519 -----IGLAGPNSVLPERALLVGSLEYQLPFTRTLGSAGVFDHMDGDAANFK 565
Db 493 GYVLSNTLTVPVQPSLLGKQASVAPFVGADVGLKNSHPDARTIRMA----- 539
Qy 566 RMLKLHSGSLGVNRFSPAPFSF 588
Db 540 -----GLAAGVREDLPYARMSF 556

RESULT 31
CARB_VIBCH STANDARD; PRT; 1076 AA.
AC O9KPH9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
CN CARB OR VC2389.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
```

```
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O -> 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -I- COFACTOR: Binds three manganese ions (By similarity).
CC -I- PATHWAY: Arginine biosynthesis.
CC -I- SUBUNIT: Pyrimidine biosynthesis; first step.
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -I- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004309; AAF95532.1; -.
CC HSSP; P00968; 1A9X.
CC TIGR; VC2389; -.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS-like.
CC Pfam; PF00289; CPasease_L_chain; 2.
CC Pfam; PF02786; CPasease_L_D2; 2.
CC Pfam; PF02787; CPasease_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPASE.
CC PROSITE; PS00866; CPASE_1; 2.
CC PROSITE; PS00867; CPASE_2; 2.
CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
CC ATP-binding; Manganese; Complete proteome.
CC DOMAIN 1 403
CC DOMAIN 404 553
CC DOMAIN 554 935
CC DOMAIN 936 1076
CC REPEAT 1 553
CC REPEAT 554 1076
CC NP_BIND 153 210
CC NP_BIND 303 354
CC METAL 285 285
CC METAL 299 299
CC METAL 301 301
CC METAL 828 828
CC METAL 840 840
CC SEQUENCE 1076 AA; 117911 MW; 82EF8623B247E4E1 CRC64;

Query Match
Best Local Similarity 3.4%; Score 109; DB 1; Length 1076;
Matches 138; Conservative 80; Mismatches 260; Indels 184; Gaps 31;

Qy 27 SENKAAGFA-----LFFKNKS-----POTESVKLKPFP---PV--RIDTQD 61
Db 474 AEVKAAGGAGLNVQDVLVRKMKRGFSDFARLSKLLGVSENIIRLRDQYNIHPVYKRVDTCA 533
Qy 62 SEIKDMVEEHLPLITQOQEEVLDKEOTGFLAAEAP-DNVKTMLRSGK-----YF 109
Db 534 AEFKS-----DTAYMYSTYDEE-----CEANPTDKDKIMVLCGGPNRIGQGLEFDYC 580
Qy 110 SSKVSLT-EKDGYATVHTPGTRTKIANVGVA-----ILGDTLS-----DCNLAE 153
```

```
Db 581 CVHAALALREDCYETIMVNCNPETVSTDYDTSRLYFEPVLTLEDVLAIVRVEKPKGVIVO 640
Qy 154 Y-----YRNLENNQOPV-GSDFDQDQSWENSKTSVLGAVTRKCYPLAKLNGNTRAAVNP 205
Db 641 YGGQTPPLKARALEAAGVPVIGTSDAIDRAEDRERFOQAQVORGLK-----QP 689
Qy 206 DTATVD-LNVVDSGRPIAFG-----DFEITGQRYPEQIVSG-----242
Db 690 DNATVTAIEQAIEKSREIGFPLVVRPSYVLGGRAMEIVYDEODLRRYFNEAVSVNESPV 749
Qy 243 -LARE-QPCTPYDLDLIL-----FQALEQNGHYSGASVQADFRLQGDVPPKVS 292
Db 750 LLDRELDDATEVDVDAICDGERVWJGGIMEHIEQAGVHSGDSACS-----LPAYTL 800
Qy 293 VTEVRHKLETGIRLDSYGLGGKAYDYNNLFNKGYIGSVVMDMDKVTETTLAAGISOPR 352
Db 801 SOEIQDKREQVEKLAFLGVLGRGLMNIOFPAVDNEVYLIEV-----NPR 844
Qy 353 NYRGNWTS-----NVSYNRSTONLEKRAFSGGI-----WYVRDRA-----GIDA 393
Db 845 AARTVPFVSKATGAPLAKIAARVMVVGOTLEQOGFTKEIIPVYSVKEVVLFPNKPFGVDP 904
Qy 394 RLGAFLAEGRKIPGSDIDLGNHATMTASWKROLLNNVLPENGVHLDCKIGITLTGF 453
Db 905 LLGPEMRSTG-----EVMGVGATFAEYAKAEELGCGSVYPEGGRAL-----LSVREGDK 953
Qy 454 LSSTALIRTSARAGYFTEPKKLGTFIIRGOAGVTVARDNADVPGLMFRSGGASSVRG 513
Db 954 QRVDLASKLVGLQDLATH-----GTAVILGEAGINPLVN-KYHEG-----RPHILDRKN 1006
Qy 514 YELDSI-GLAGPNSVLPERALLVGSLEYQLPFTTTLSCAVFHDH-----GDAAANFKRML 569
Db 1007 HEYTVIVNTASGQAIEDSKVLRRGALAHKVNYYTTTLNAATCATCSHTADAKASVTSQVE 1066
Qy 570 KH 571
Db 1067 LH 1068
```

RESULT 32

```
MYFC_YEREN
ID MYFC_YEREN STANDARD; PRT; 841 AA.
AC P33408;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane usher protein myc precursor.
GN MYFC.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W1024 / Serotype O:9;
RX MEDLINE=94018646; PubMed=8105362;
RA Irlarte M., Vanooteghem J.-C., Delor I., Diaz R., Knutton S.,
RA Cornells G.R.;
RT "The myc fibrillae of Yersinia enterocolitica.";
RL MOL. Microbiol. 9:507-520(1993).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE MYFA
CC FIMBRIAL SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: 221953; CA79953.1;
DR PIR: S39365; S39365;
DR InterPro: IPR000015; Fimb_usher.
DR Pfam: PF00577; Usher; 1.
DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 841 OUTER MEMBRANE USHER PROTEIN MYFC.
FT DISULFID 817 840 POTENTIAL.
SQ SEQUENCE 841 AA; 93414 MW; B6D207F75161D0D0 CRC64;

Query Match 3.4%; Score 108.5; DB 1; Length 841;
Best Local Similarity 18.7%; Pred. No. 17;
Matches 124; Conservative 82; Mismatches 232; Indels 225; Gaps 32;

Qy 16 FPHAYAPAAADLSENKAA-----GFALFKNKSPDTE-----SVLKPKFPVRIDTQ 60
Db 111 YPELLPPEKNIQQDKQASQCVNLAAIPOASEFEFYAMRLVLNIPQVALRPKDEIPVERW 170
Qy 61 DSEIKDMVEEHLPLITOOQEEVLDKEQTGFLEAEAPDNVKTMLRSKGYFSSKVSLEKDG 120
Db 171 DQGI-----TAFLLNYMANSSTTTTQTG-----EQOS 198
Qy 121 AVTHITPGPRTKIANVGVAIILGDLSDGNLAEEYRNALENWQOPVGSDFDQDQSWENSKT 180
Db 199 SHYQLYFG-----FNIG-----AWRIINA-TSNQ-----SGNSACKWQSSYI 236
Qy 181 SVLGAVTRKGYPLAKLNGNTRAAVNPDTATVD-LNVVDSGRPIAFGDFEITGQRYPEQI 239
Db 237 ----YATRGLYRL-----KSRVTLGOSYTPGDFDPSVGVMLGDDNNMLPSSQDFIPV 288
Qy 240 VSGLARFO-----PGTPYDL-DLLLDFOQ-----ALEONGHYSGAS 274
Db 289 VRGIARSOARVEVRONGYLLYSTVYVSPG-PFELTDILPSHNSGDLHVTYLESNG----- 341
Qy 275 VOADFRLQGDGRVPVKVSVTEVRHKLETGI-----RLDSE-----YGLGK 316
Db 342 -----TTQQFTVPYSPVPAIRLKRGLRYNLNMGARYRANVDVETPTAQATVAVGLP-- 393
Qy 317 IAYDYNNLFNKGYIGSVVMDMDKYETTLAAGISQPNRYGNVYTSVNSYRSTTONLEKR 376
Db 394 -----WNL--TAFVGG-QWSPHYQATTAGMGV-----MLGDYGALSSSITQATSEYRQOQ 440
Qy 377 AFSGGIWYVRDAGIDARLGAFFLAEGRKIPGSDIDLGNHATMTASWKROLLNNVLP 436
Db 441 PVKGQVWEVR-----YNKTLQASDTSFVSNVSNQYSTADFSN--LSDVLQS 483
Qy 437 ENGHYLDCKIGTTLGTFLSSTALIRTSARAGYFTEPKKLGTFIIRGOAGVTVARDNAD 496
Db 484 YRRHDYSRR-----DWHNSNLRNTHVVVG-----QPLQGF-----GYLNLNWSRON-- 525
Qy 497 VPSGLMFRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLEYQLPFT-----R 547
Db 526 -----YRDAPASSSMGVQY-----SFNIGNLYCSLDWTQYRGNDQR 563
Qy 548 TILSGAVFHDMDGDAANFRKMLKHCGLGVRFWFSPLAFPSFDIAYGHS--DKKIRHHISLG 606
Db 564 LLSLSVSMPLGRDRTYAAAYRMTSSDNKSDHEMS-----LYCHAFDNRLSNVNRQT 614
Qy 607 TRF 609
Db 615 ERY 617

RESULT 33
GTFC_STRMU
ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
```



```

CC -1- SUBUNIT: DNA-BINDING HOMOTRIMER IN STRESSED OR HEAT SHOCKED CELLS,
CC OTHERWISE FOUND AS A HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC DURING NORMAL GROWTH AND MOVES
CC TO THE NUCLEUS UPON ACTIVATION.
CC -1- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X61754; CA443893.1; ..
CC PIR: B40583; B40583.
CC HSP: P22121; 2HTS.
CC TRANSFAC: T00972; ..
CC MGD: MGI:96239; Hsf2.
CC InterPro: IPR00232; HSF_DNA_bind.
CC InterPro: IPR002341; HSF_ETS.
CC Pfam: PF00447; HSF_DNA_bind; 1.
CC PRINTS: PR00056; HSFDOMAIN.
CC ProDom: PD001788; HSF_DNA_bind; 1.
CC SMART: SM00415; HSF; 1.
CC PROSITE: PS00434; HSF_DOMAIN; 1.
CC Transcription regulation; Nuclear protein; DNA-binding; Activator;
CC phosphorylation; Heat shock; Multigene family.
CC DNA_BIND 7 112 BY SIMILARITY.
CC DOMAIN 119 192 HYDROPHOBIC REPEAT HR-A/B.
CC FT DOMAIN 359 384 HYDROPHOBIC REPEAT HR-C.
CC FT DOMAIN 517 517 AA; 58155 MW; 098965BBCDF4E1B0 CRC64;
CC SQ SEQUENCE 517 AA; 58155 MW; 098965BBCDF4E1B0 CRC64;
CC -----
Query Match 3.48; Score 108; DB 1; Length 517;
Best Local Similarity 21.38; Pred. No. 9.4;
Matches 64; Conservative 45; Mismatches 91; Indels 100; Gaps 13;
Qy 44 TESVKLPKFPVRIDTQSEIKDWEEHLPLITQOOEVL-----DKE-- 86
Db 233 TEGLSKRERISDDIIIVDTDDNDENIPVPIETNEDVVDSSNOYPDIVIVEDNEDE 292
Qy 87 -----QTGFLAEAPDNVKT-----MLRSKYFSSKSVLTKDKGAYVHTPGPRT- 132
Db 293 YAPVIQSQEQSEAPREPLRVGSAGSSPLMSSAVQLNGSSLSITSED-----PVTM 342
Qy 133 --KTANVGVAILDILSDGNLAAYRN---ALENKQOPVGS-----DFDQDSWE 176
Db 343 MDSILNDNINILGKV---ELLDYLDSDICSLEDFQAMLSGRQFSIDPDLLVDSENKGL 398
Qy 177 NSKTSVLGAVTRKG-----YPLAKL--GNTRAAVNPDTATVDLNVVVD 218
Db 399 ATKSSVVQHVSEGRKSKSPDKQLIOYAFPLAFLDGNASALEQGSSTTASSEVPSV 458
Qy 219 GRPIAFGDFEITGTQRYPEQIVSGLARFPQT-----PYDLDL-LLD 259
Db 459 DKPIEVD--ELLDSSLDPEPTQSKLVRLPLEUTEAEASEATFLYLCELAPALDSMDPLD 516
RESULT 35
FLA3_CAMJE STANDARD; PRT; 574 AA.
AC Q46113; 001673;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin A.
GN FLAA OR FLAI.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_Taxid=197;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ATCC 43431 / TGH 9011 / Serotype LIO36;
RA Khawaja R., Neote K., Bingham H.L., Penner J.L., Chan V.L.;
RT "Cloning and sequence analysis of the flagellin gene of Campylobacter
RL Curr. Microbiol.", 24:213-221(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43431 / TGH 9011;
RA Chan V.L., Bingham H.L.;
RT "Organization and sequence of the flagellin genes of Campylobacter
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M74578; AAA17046.1; ..
CC EMBL: Z29327; CAAB2523.1; ..
CC InterPro: IPR001492; FlagellinN.
CC InterPro: IPR001029; Flagellin_C.
CC Pfam: PF00669; Flagellin_N; 1.
CC Pfam: PF00700; Flagellin_C; 1.
CC PRINTS: PR00207; FLAGELLIN.
CC ProDom: PD000316; Flagellin_C; 1.
CC Flagella; Repeat.
CC INIT_MET 0 0 BY SIMILARITY.
CC REPEAT 404 408 1.
CC REPEAT 410 414 2.
CC REPEAT 446 449 3.
CC CONFLICT 512 514 VTS -> LOV (IN REF. 1).
CC CONFLICT 532 532 Q -> T (IN REF. 1).
CC CONFLICT 545 545 Y -> F (IN REF. 1).
CC CONFLICT 548 548 A -> Y (IN REF. 1).
CC CONFLICT 560 560 A -> S (IN REF. 1).
CC CONFLICT 564 564 S -> A (IN REF. 1).
CC CONFLICT 571 571 R -> K (IN REF. 1).
CC SEQUENCE 574 AA; 59483 MW; 7696ABD54FA36B85 CRC64;
CC -----
Query Match 3.48; Score 108; DB 1; Length 574;
Best Local Similarity 20.98; Pred. No. 11;
Matches 106; Conservative 68; Mismatches 186; Indels 148; Gaps 27;
Qy 138 GVAILGDTLSDGNLAAYRNALENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLG 197
Db 46 GMAIKDSLRSQANT---LGQAISNGNDALGILQTADKAMDEQLKILDTIKTKATQAAQDG 102
Qy 198 ---NTRAAVNPDTATVDLNVVVDGSRPIAFGDFEITGTQRYPEQIVSGLARFPQPTPYDL 254
Db 103 QSLKTRTMLO-----ADINRLMEELDNTA-----NTTSPNGKQLLSC----- 139
Qy 255 DLLLLDFOQALQNGHYSGASVQADFDRLQGDVRPVKSVTVKRRHKLETGIRLDSYGLG 314
Db 140 ---NFINQEFQIGASSNOTVKASIGATQSS---KIGLT-----RFETGSRI---SVG 182
Qy 315 GKIAVDYNNLFPNKGIVGVSVDMDK--YETTLAAGISOPRNVRYGNWTSNVSYNRSTQN 372
Db 183 GEVOFTL-----KNYNGIDDFKFKQKVIISTSVGTGLG-----ALADE 219
Qy 373 LEKRAFSGGI---WYVRDRAGIDARLGA---EFLAEGRRKIPGSDIDLGNSHATMLTSMK 426
Db 220 INKNADKTGVRAATFVETRGCAVVRAGATSDDFAINGVKIKVDYKGDGANGALVSA--- 276
Qy 427 ROLLNNVLH-----PENGHYL-----DGKIGTTLGTFLLST-----AL 459

```

```
Db 277 ----INSVKDTTGVASIDENGKLLLTSGREGIKIEGNI--GAFINPNMLENYGRSL 331
Qy 460 IRTSARAGYFFTPENKK---LGTFTIRGQAGTYVARD-----NADVPGLMFRS----- 505
Db 332 VANDCK-DLISGTLNLSAIGFTGNWISOASVLSRESKQIDANDAMGFNSANKGNIL 390
Qy 506 GASSVRGYELDISGLAGNSVLPERALLVGS-LEYQLPFTRTLGS-----GAVFH-- 555
Db 391 GGYSSVAY-MSSTGSGSSGS-----GFSVSGSKNYSTGFANTIAISAALSAYVNV 444
Qy 556 -----DMGDAANFRKMLKHGSLGVR 578
Db 445 AGSGFSSGSLNSQFATMTKTSAGNTLGVK 472

RESULT 36
PRIC_ERWCH STANDARD; PRT; 479 AA.
AC P16317;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Secreted protease C precursor (EC 3.4.24.-) (Proc).
GN PRIC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=B374;
RX MEDLINE=91009140; PubMed=2211614;
RA Deleplaire P., Wandersman C.;
RT "Protein secretion in Gram-negative bacteria. The extracellular
RT metalloprotease B from Erwinia chrysanthemi contains a C-terminal
RT secretion signal analogous to that of Escherichia coli alpha-
RT hemolysin."
RL J. Biol. Chem. 265:17118-17125(1990).
RN [2]
RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=B374;
RX MEDLINE=8925387; PubMed=2722818;
RA Deleplaire P., Wandersman C.;
RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are
RT synthesized and secreted as zymogens without a signal peptide."
RL J. Biol. Chem. 264:9083-9089(1989).
CC -I- COPACTOR: THE DIVALENTS CATIONS MG(2+) AND CA(2+) INCREASE
CC PROTEASE ACTIVITY SIGNIFICANTLY ABOVE THE CONTROL VALUE.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR
CC SECRETION OF THIS METALLOPROTEASE.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M59229; AAA24860.1; .
CC DR EMBL; J04736; AAA24862.1; .
CC DR EMBL; M60395; AAA63638.1; .
CC DR PIR; A38307; A38307.
CC DR HSSP; P23694; 1AF0.
CC DR MEROPS; M10.054; .
CC DR InterPro; IPR001343; Hemlysn_Ca_bind.
CC DR InterPro; IPR000130; Zn_Mrptptdse.
CC DR Pfam; PF00353; hemolysinCabin; 3.
CC DR PRINTS; PR00313; CABNDNCRPT.
CC DR SMART; SM00235; znmC; 1.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
```

```
DR PROSITE; PS00330; HEMOLYSIN_CALCITUM; 1.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Repeat.
FT PROPEP 1 17
FT CHAIN 18 479 SECRETED PROTEASE C.
FT METAL 188 188 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 189 189 BY SIMILARITY.
FT METAL 192 192 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 345 386 5 X REPEATS, GLY-RICH.
FT REPEAT 345 350 1.
FT REPEAT 354 359 2.
FT REPEAT 363 368 3.
FT REPEAT 372 377 4.
FT REPEAT 381 386 5.
SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;

Query Match 3.4%; Score 107.5; DB 1; Length 479;
Best Local Similarity 22.4%; Pred. No. 9.1;
Matches 104; Conservative 58; Mismatches 168; Indels 135; Gaps 27;

Qy 88 TGFVLAEEAPDNVKTMLRSKGYFS-SKVSLTEKDGAVTVHITPGPRTKIANVGVAIIGDL 146
Db 92 TGFVFNAAEQTEQAKLSQSDVANLTFTEVTGNKSNITFGNTRDAS----- 141
Qy 147 SDGNL-----AEY---YRNALENMQQPVGSDFOQDSWENSKTSVLGAVTRKGYPLAKL 196
Db 142 --GNLDYGTQAYAYYPGNVQGAGSSWY-----NYNQSNIIRNPGSEYGRQTF---THEI 190
Qy 197 GNTRAAVNPDTATVDLNVVDSGRPIAFQDFEITQYRPEQIVSGLARFQGTPTDIDL 256
Db 191 GHALGLAHFG-----EYN--AGEGDP--SYND--AVYAEDSYQFSIMSNGENETGADY--- 238
Qy 257 LLDFQALAEQNGHYSYGASVQAD---FDRLOGDRVPVKVSTEVKRRKLETGIRLSEYGL 313
Db 239 -----NGHYCGAPMIDIAIQRLYGANWTR-----TG---DSVYGF 273
Qy 314 GKKIADYVNL--FNKGYIGSVVMDMKYETTLAAGISQPRNRYGNWTSNVYNRSTQ 371
Db 274 NSNTORDEFYATDSSKALIFS-VMD-----AGGDTDFSG-----YNNQRI 315
Qy 372 NLEKRAFSGGIWYVRDRAGIDARLCAEFLAERKRI-----PGSDIDLGNSHATWLTASW 425
Db 316 NLNEGSFS-----DVGLKGNVS---IAHGVTTIENAGSGNDILVGNSSADNLIQGG- 364
Qy 426 KQLLNVLHPENG-HYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKK-----LG 478
Db 365 ---ACNDVLYCGAGADTLYG--GAGEDTFVYVYSG--QDSTVAAYDWIADFQKIDKIDLS 417
Qy 479 TPIIRGQAGYT-----VARDNADVPGLMFRSGGASSV 511
Db 418 AFRNEGQLSFVQDQFTGKQGEVNLQWDAANSITNLWLHEAGHSSV 462

RESULT 37
PRIC_PSESR STANDARD; PRT; 587 AA.
ID PRIC_PSESR
AC P42790;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pseudomonaspepsin precursor (EC 3.4.21.100) (Pepstatin-Insensitive
DE DE carboxyl proteinase).
GN PCP.
OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33067;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
RX MEDLINE=95014496; PubMed=7929375;
RA Oda K., Takahashi T., Tokuda Y., Shibano Y., Takahashi S.;
RT "Cloning, nucleotide sequence, and expression of an isovaleryl
RT pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
RT 101.";
```

J. Biol. Chem. 269:26518-26524 (1994).

[2] SEQUENCE OF 216-585.

RA MEDLINE=96157710; PubMed=8576087;

RA Hayashi K., Izu H., Oda K., Fukuhara K.-I., Matsuo M., Takano R.,

RA Hara S.;

RT "The primary structure of pepstatin-insensitive carboxyl proteinase

RT produced by *Pseudomonas* sp. No. 101.,"

RA J. Biochem. 118:738-744 (1995).

[3] MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE

RP SITES.

RA MEDLINE=99419069; PubMed=10488127;

RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;

RT "Identification of catalytic residues of pepstatin-insensitive

RT carboxyl proteinases from prokaryotes by site-directed mutagenesis.,"

RL J. Biol. Chem. 274:27815-27822 (1999).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-

CC Glu-1-Ala-14, 15-Leu-1-Tyr-16 and 25-Phe-1-Tyr-26 and angiotensin

CC I at 4-Tyr-1-Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-

CC Phe-1-Phe(NO₂)-Arg-Leu.

CC -1- SUBCELLULAR LOCATION: Periplasmic.

CC -1- PTM: AUTOCATALYTICALLY PROCESSED.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: D37970; BAA07188.1; -

DR MEROPS: S53.001; -

KW Hydrolase; Protease; ?

FT SIGNAL 1

FT PROPEP 215 REMOVED IN MATURE FORM.

FT CHAIN 216 585 PSEUDOMONAPEPSIN.

FT PROPEP 586 587 REMOVED IN MATURE FORM.

FT ACT_SITE 385 385 PROBABLE.

FT ACT_SITE 502 502 NUCLEOPHILE (BY SIMILARITY).

FT ACT_SITE 543 543 PROBABLE.

FT DISULFID 352 391

SQ SEQUENCE 587 AA; 61072 MW; E193D8B2C25829A CRC64;

Query Match 3.48; Score 107; DB 1; Length 587;

Best Local Similarity 20.48; Pred. No. 13;

Matches 137; Conservative 69; Mismatches 226; Indels 238; Gaps 34;

Qy 21 APAADLSNKAAGFALPKNKPDPSTESVKLPKFPVRIDTQDSSEIKDMVEEHLPLITQQE 80

Db 60 APLELAGETAHIVV-----SLKLR-----DEAQLKLAQAVNQGNAGFG 101

Qy 81 EVLDKEQTGFLEAPDNVKTMLRSKGFSSKVLTEKDGAYTVHTPGPRTKIANGVGA 140

Db 102 KFLKRRQ--FLSQFAP-----TEAQVAVVAHLKRGFNHIVPN-RLLISADGSA 150

Qy 141 ILGDLSDGN--LAERYNALENWQOPVGSDFDQDSWENSKTSVLGA--VTRKGYPLAKL 196

Db 151 --GAVKAAPNPLVRYQLNGKAGYANTAPAQVQDGLGE-IVGVLGLQNVTR-AHPMLKV 206

Qy 197 GNTRAAVNPDTAT-----VDLNVVVD--SGRPIA-----FGDFEITGTQRYPEQIVSGLAR 245

Db 207 GERSNAKTLAGTAKGHNPTEFFPIYDASSAPTANTVTGIIITGGVSQ----- 255

Qy 246 FOPGTPYDLDLLDFQQAQLHSGVSGASVQADFRLQGRVPVKVSYTEVYKRLKETGI 305

Db 256 -----TLQDLQOFTSANG-----LASVNTQTQITG- 280

Qy 306 RLDSEYGLGGKIADYNNFLNFKNGIGSVVWDMKYEYTTLAG--ISQPRNTRGNWTS-- 361

Db 281 ---SSNG-----DY-----SDDQOQGGENDLDSQSVGSAGAVQQLLFYMQDSASGN 326

Qy 362 ---NVSYNRSTQNLKRAFGSGGIWYVRDRAGIDARLGAEE-----FLAEGRKIPGSDID 412

Db 327 TGLTQAFNQAVSDNVAK-VINVSLGWCEADANADGTLOAEDRIATRAAQGTSTSVSSGD 385

Qy 413 LGNSHATMLTASWKROLNNVLPHPENGHYLDGKI-----GTTTGTFLUSS 456

Db 386 EG-----VYECNNRGYPDGSYSTSVSWPASSPNVIAVGTTTLYT----- 423

Qy 457 TALIRTSARAGYFTTPENKKLGTIIIRCOAGTYVARONADVPVSGLMFSPSGCASSVRGYE- 515

Db 424 -----TSAGA-----YSNETVMNEGLD-SNGKLWATGGGYSV--YES 457

Qy 516 ---LDSTGLAGPNGSVLP-----ERALLVGSLEYQLPFTRTLSG----- 551

Db 458 KPSQSVSVSTPGRRLLPDISFDAAGTGALIVYVGOIQIGGTSLASPIFVGLWARLOS 517

Qy 552 AVFHDMDGAANF-----KRMKLKH-----GSGLGVRWFSPFLAPSFDA--- 591

Db 518 ANSNLGFPAASFYSAISSTPSLVHDVKSGNGGYGGYNAGTGWDYPTGWSLDIAKLS 577

Qy 592 -----YGH 594

Db 578 AYIRSNRNGFGH 587

RESULT 38

K6PP_RABIT

ID K6PP_RABIT STANDARD; PRT; 791 AA.

AC P47859;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 6-phosphofructokinase, type C (EC 2.7.1.11) (Phosphofructokinase

DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme C) (PFK-C).

GN PFKP.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

[1]

RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Brain;

RX MEDLINE=94164929; PubMed=8119919;

RA Li Y, Valaitis A.P., Latshaw S.P., Kwiatkowska D., Tripathi R.L.,

RA Campbell M.C., Kemp R.G.;

RT "Structure and expression of the cDNA for the C isozyme of

RT phosphofructo-1-kinase from rabbit brain.,"

RL J. Biol. Chem. 269:5781-5787 (1994).

CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-

CC fructose 1,6-bisphosphate.

CC -1- COFACTOR: MAGNESIUM.

CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR

CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.

CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN

CC SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: U01154; AAA17707.1; -

DR HSHP; P00512; 3PFK.

DR InterPro: IPR000023; Ppfuckinase.

DR Pfam: PF00365; PFK; 2.

DR PRINTS: PR00476; PPFRCCKINASE.

DR ProDom: PD000707; Ppfuckinase; 2.

DR PROSITE: PS00433; PHOSPHOFRUCTOKINASE; 2.

KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
SQ SEQUENCE 791 AA: 86349 MW: 3C10A36F229FD8E8 CRC64;

Query Match 3.4%; Score 107; DB 1; Length 791;
Best Local Similarity 18.9%; Pred. No. 20;
Matches 118; Conservative 77; Mismatches 265; Indels 164; Gaps 25;

QY 81 EYLDKEQGTGLAEAPDNVKTMLRSKGVFSSKVSJTEKDGAVTVHITGCP----- 130
DB 196 EVIDAITTAQSHQTFVLEVMGRICGGLVALVSALA--CGADWFLPSPPEEGWEOMC 253
QY 131 -----RTKIANVGVAIGDILSDGNLAIEYRNALENMQQPVGSDFDODSWNSKTSVLG 184
DB 254 VKLSENRAQKKRLNIIIAE-----GAIDLNRPTS-----EKIKELV-- 292
QY 185 AVTRKGYPLAKGNTRAAVNPDTATVDLVNVVDSRPTAFGDFEITGQRYPEQIVSGLA 244
DB 293 -VTQLGY-----DTRVTILGHVQGGTPSA---FDRILASRMGVEAVLALL 334
QY 245 RFQPCGTP-----YDLDDLDD---FQQALEONGH-----Y 270
DB 335 EATPETAQVSVLSNHNHVAHVLPLVECVQMTQEVQKAMDERREFKDAVQLRGSRFENNLT 394
QY 271 SGASVQADFDRLQGRVPVKVSVTEVKRHKLETKIRLSEYGL--GGKI--AYDYNNLFN 326
DB 395 KRLAIKLPDDKIQKNCNVAVINVCAPAGMNAVRSVRVGIADGKMFAYVDGDFGFA 454
QY 327 KYIGSVVW--DMKVETTLAIGISOPRNVGNWTSVNSYNSTTONLEKRAFSGGIWV 385
DB 455 KQIKIRMGVGGVGTGGGSGILGTRILPKGYLEIATQIR--THNINAILIIG----- 507
QY 386 RDRAGIDARGLAEFLAERK-----IPGSDIDLGNSHATMLTASWK 426
DB 508 -----GFEAYLGLLELSAAREKHEEFCVPMVMPVATVNSNVPDGSFGADTALNTIDTC 563
QY 427 ROLLNNVLPENHYLDGKIGTTLGTLFSLSTALIRTSARAGYFFTPENKKLGTFTIIR--- 483
DB 564 DRIKQSASCTKRRVFTIETMGYCG-YLANMGGLAAGADAAYIFEE-----PFDIRDLQ 616
QY 484 GQAGTYVARDNADVPGLMFRSGGASSVRGYELDSI-GLAGPNG-SVLPERALLVGSLEY 541
DB 617 SNVEHLTEKMKTTIQRGLVLRNESCS--ENYTTDFIYQLYSEGRGVDFCRKNVLGHMQQ 674
QY 542 ---QLPFTTLSCAVFHDMDGDAANFKRMLKKGSGLGVW-----PS 581
DB 675 GCAPSPFDNFCTKI-----SARAMQWITTKLESFGKGRFVSDDSCIVLGISKRNVLQ 730
QY 582 PLAPFSFDIAYGHSDKKIRWHISL 605
DB 731 PVAEKNETDFEHRIPKEQWMLKL 754

RESULT 39
AFAC_ECOLI STANDARD; PRT: 859 AA.
AC P53517;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane usher protein afac precursor.
GN AFAC.
OS Escherichia coli.
OG Plasmid p11055.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A30;
RX MEDLINE=95095929; PubMed=8002584;
RA Garcia M.-I., Labigne A., le Bouguenec C.L.;
RT "Nucleotide sequence of the afimbrial-adhesin-encoding afac-3 gene

cluster and its translocation via flanking IS1 insertion sequences.";
J. Bacteriol. 176:7601-7613(1994).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF AFA-III AFIMBRIAL
CC ADHESIN SUBUNITS ACROSS THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT PROTEIN FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76688; CAA54117.1; -;
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal; Plasmid.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 859 OUTER MEMBRANE USHER PROTEIN AFAC.
SQ SEQUENCE 859 AA: 93805 MW: 1FF7FE3715EBF93 CRC64;
Query Match 3.4%; Score 107; DB 1; Length 859;
Best Local Similarity 20.1%; Pred. No. 22;
Matches 135; Conservative 79; Mismatches 189; Indels 270; Gaps 38;
QY 57 IDTQSEIKDMVEEHLPLT-----TQOQEEVLDDKQGTGLAEAPDNVKTMLRSKGYFS- 110
DB 278 MSSDESMPYNLREFAPVVRGIARTQARIEV---RQGYL-----IQSQTVAFGAFAL 327
QY 111 -----SKVSLTEKDGAYTVHITPGPRTKIANVGVAIGDILSDGNL-----ABY 154
DB 328 TDLPTVGTSGSDIQVTVLESDDGTAQVTFPTTPAIA-----LREGYLKYNVTAQ 377
QY 155 YRNALENMQQPVGSDFDODSWNSKTSVLGAVTRK-GYP--LAKIGNTRAAVNPDTATVD 211
DB 378 YRSS-----DDAVEHTS---LGQVTAMGYLPWGLTVYGGLOGADDYSAALG 421
QY 212 LN-----VVVDS-----GRPIAFGD-----FEITGQRYPEQIVSGLARF 246
DB 422 LGWSLGRGLGAVSLDTHSRGQOKGHDYETGDTWRINKSELTCT----- 467
QY 247 QPCTPYDLDDLDFQOALEQ---NGHYSASVQADFDRLQGRVPVKVSVTEVKRHKLET 303
DB 468 -----SFTAASYQSSDGYHTLPDV---LDTWRDDRYAYRHTENRSRR----T 508
QY 304 GIRLDSYGLGKIAV---DYNNLFNKGYIG---SVVWMDKYETTLLAAGISQPNRYR 355
DB 509 TULSQSLQGWYGVGLNGSRDEYRDRPHRDYEGASYTSWN---NISLVNWSNRNRSG 564
QY 356 GNY-----WTSNVSYNRSTTONLEK-----RAFSG 381
DB 565 GYYGWSRTEDSVSMMSVPLGRWFGGADNDISTTAQMRSTGQDTRYEAGLNGRAFDRR 624
QY 382 I-WYVRDRAGIDARGLAEFLAERKIPGSDIDLGNSHA--TWLTASW----- 425
DB 625 LYWDVRE-----QMVPGSE-----SHADTSRLNTWTGYTGELTCWYSY 663
QY 426 ---KRQLL-----NNVLHPENHYLDGKIGTTL-----GTLSSTALLRTSAR--- 465
DB 664 STMQLNLNMGSGSVVAHSE-CVTFGQRTGDTVALIAAPGVSGASVGGHPGVTRDFRGYT 722
QY 466 -AGYFPTPENKKLGTFIIRGQAGYTVARDNADVP-----SGLMFRSGGASSVRGYE 515
DB 723 LAGY-ASPYQENVLTL-----DPTTFPEDAEVPTDSRVVPTKGAVVRAGRPTRVGGRA 775
QY 516 LDSIGLAGPNSVLP-----ERALLVGSLEYQLPFTRLTSGAVFHDMDAANFKR 566
DB 776 L--VSLARODGTPLPFGAVTVVEGERGQAAG-----SAGVVGDRGEVLYS--- 818


```
Qy 567 MCLKHGSLGVWR 579
    || | : |
Db 819 -GLRESGKLKAQW 830

RESULT 40
E2BI_PYRAB
ID E2BI_PYRAB STANDARD; PRT; 356 AA.
AC Q9UZ16;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative translation initiation factor eIF-2B subunit 1 (eIF-2B
DE GDP-GTP exchange factor).
GN AIF-2BI OR PAB2444.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE EXCHANGE OF INITIATION FACTOR 2-BOUND GDP
CC FOR GTP (BY SIMILARITY).
CC -!- SUBUNIT: COMPLEX OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE EIF-2B ALPHA/BETA/DELTA SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ248287; CAB50246.1;
CC InterPro; IPR005251; AIF-2BI_fam.
CC InterPro; IPR000649; IF-2B.
CC Pfam; PF01008; IF-2B; 1.
CC TIGRFAMs; TIGR00512; aif-2BI_fam; 1.
CC TIGRFAMs; TIGR00524; eif-2B_rei; 1.
CC Hypothetical protein: Initiation factor; Protein biosynthesis;
CC Complete proteome.
CC SEQUENCE 356 AA; 39548 MW; 4273C0A06C5B7513 CRC64;
```

Search completed: November 9, 2002, 01:19:27
Job time : 30 secs

```
Db 265 AVLAR-EHGIPFFAVAPLSSIDMSLKGKPEIERSPEVLTCGGCRIAPD----- 315
Qy 287 VPVKVSVTEVKRHKLETGIRLD 308
    ||| : | | | |
Db 316 VPVYNPAFDVTPHKYVTGIITD 337
```

```
Query Match 3.4%; Score 106.5; DB 1; Length 356;
Best Local Similarity 23.6%; Pred. No. 7;
Matches 76; Conservative 49; Mismatches 120; Indels 77; Gaps 17;

Qy 21 APAADLSENKAAGFALF----KNKSPDT-----ESVKLKPFPPVRIDTQDSEIKDM 67
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 APA--IGAAAFGLALYAEYSKAKSDFMDFYKAYETLNTPTAVNLFALNRIKKL 116
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 68 VEEHLPLITQQOQEEVLDEKQTFLAEE---APDNVKTMLRSKGYFSSKSVLTKDGAYTV 124
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 VEEHL-----EDPLDEIKSLIVNEAKIADDEDVEANLR-MGHYGAEVL---PEGNLLT 165
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 125 HITPGPRTKIANVGVAIIIGDL-----SDGNLAEEYRNALENWQPPVGSDFODSWENSKT 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 HCNAG---SLATVHLGTGVAVVRVNMHKDGLSLKLWLDE---TRPVQCARLSAWEYS-- 216
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVNVVDSGRPIAFGDFEITGTRYPEQIV 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 -----YDGLNVKLIADNAAAFVMOQGLVDA-IIVGADRIVANGDF-----ANKIGTYML 264
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 241 SGLARFPQPTPY-----DLDLILLDFQQALEQNG-----HYSCASVQADFDRLOGDR 286
    : ||| : | | : | | : | | : | | : | | : | | : | | : | | : | |
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 01:23:59 ; Search time 1941 Seconds
(without alignments)
5081.428 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKIRWHISLQTRF 609

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO-POOL/US09857669/runat_05112002_110008_7737/app_query.fasta_1.775
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669 -ACGN_1_1045_@runat_05112002_110008_7737 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRLADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	141.5	4.5	792	14	BQ506223
c 2	129	4.1	607	10	AV557254
c 3	125	3.9	649	17	BH396320
c 4	121	3.8	503	10	BE459389
c 5	119	3.8	494	10	AW191135
c 6	119	3.8	2776	11	AK002711
c 7	117	3.7	1829	11	AY109520
c 8	111	3.5	588	10	AW053211
c 9	109.5	3.5	503	11	AY108685
c 10	108	3.4	3045	11	AK004789
c 11	107	3.4	1077	13	BI559235
c 12	105.5	3.3	353	14	D46516
c 13	103.5	3.3	620	12	BF942509
c 14	103.5	3.3	928	14	BQ723872
c 15	103	3.2	423	9	AI363592
c 16	103	3.2	2936	11	AK011769
c 17	101.5	3.2	451	13	BM438066
c 18	101.5	3.2	1031	14	BQ876247
c 19	101.5	3.2	3095	17	BH770975
c 20	100	3.2	724	14	BQ872838
c 21	99	3.1	613	14	BQ871231
c 22	99	3.1	1171	13	BM560625
c 23	98	3.1	622	14	BQ875918
c 24	98	3.1	994	12	BF535617
c 25	97.5	3.1	4944	11	BC035446
c 26	97	3.1	630	14	BQ871709
c 27	96.5	3.0	690	14	BQ479447
c 28	96	3.0	577	14	BQ872562
c 29	96	3.0	660	14	BQ875018
c 30	96	3.0	677	14	BQ875030
c 31	96	3.0	731	14	BQ874536
c 32	96	3.0	735	14	BQ850925
c 33	96	3.0	757	14	BQ871570
c 34	96	3.0	1072	14	BM919327
c 35	95.5	3.0	963	14	BQ675465
c 36	95.5	3.0	1086	12	BG704719
c 37	95.5	3.0	2466	11	BC013324
c 38	95	3.0	578	14	BQ875852
c 39	95	3.0	581	14	BQ875483
c 40	95	3.0	622	14	BQ874935
c 41	95	3.0	659	14	BQ858721
c 42	95	3.0	660	14	BQ874292
c 43	95	3.0	662	14	BQ875986
c 44	95	3.0	669	14	BQ874486
c 45	95	3.0	671	14	BQ875563

ALIGNMENTS

RESULT 1
BQ506223/c
LOCUS EST613638 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STMGI92
DEFINITION 3' end, mRNA sequence.
ACCESSION BQ506223
VERSION BQ506223.2 GI:21922105
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 792)

AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.

TITLE Generation of a set of potato cDNA clones for microarray analyses

JOURNAL Unpublished (2002)

COMMENT On Jun 10, 2002 this sequence version replaced gl:21365092. Other_ESTS: EST613637

FEATURES Location/Qualifiers

1..792

/organism="Solanum tuberosum"

/cultivar="Kennebec or Binjfe"

/db_xref="taxon:4113"

/clone="STMGI92"

/clone_lib="Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"

/tissue_type="mixed tissues"

/lab_host="SOLR"

/note="Vector: pBluescript SK(-); site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes 'tubers, or roots.'"

BASE COUNT 236 a 220 c 159 g 177 t

ORIGIN

Alignment Scores:

Pred. No.: 3 04e-06 Length: 792

Score: 141.50 Matches: 65

Percent Similarity: 36.16% Conservative: 33

Best Local Similarity: 23.99% Mismatches: 114

Query Match: 4.46% Indels: 59

DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ506223 (1-792)

QY 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGlyLysArgAlaPheSer 379

Db 745 ACGCGGGGATTGAATACAGTCGCCATTTC-----AGGCCAAAGTGAAT 701

QY 380 GlyGlyIleTrpThrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399

Db 700 GGCACAGCCGGGAATCATATTTTCACGCTGCTGGTGGATGACAAAGGGAAATCCTATT 641

QY 400 LeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419

Db 640 ATAAGGAGCTACTACAGTACTCCACTTACT-----GCAAGTGGCAACACTCATGATGAT 587

QY 420 MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGly 439

Db 586 ATCTTACTTGCCAAACTT-----GAGACTGTC 560

QY 440 HisTrpLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459

Db 559 TATACGTGTTCTGGTCACCGCTGCTTTCAGTGTGTTTTCACATGATGATCAGGACTT 500

QY 460 -----IleArgThrSerAlaArgAlaGlyTyr 468

Db 499 CCTGTGTCACAGTCTCTAGTCTTTTCACAGAGTCAATGCTGCTAGAAAGATTC 440

QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyr 488

Db 439 GTATTAGGTCCTATGCGT-----CTTCTCTCAAGCTTCTCTGGTGCTCAT 395

QY 489 ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAla 508

Db 394 GTGCTC-----GGTAATTTTCCACCTCATGAAGCAATTTGCTGTGGTGAACC 347

QY 509 SerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVal 528

Db 346 AATAGTCTAAGAGGATATGAA-----GNAGGCACAGTT 314

QY 529 LeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThr 548

Db 313 GGCTCTGGCGGCTCTTATGCGTTGGATGTGGAGAAATTTCCCTCTCTGTATGGGGCCA 254

QY 549 LeuSerGlyAlaValPheHisaspMet----- 557

Db 253 TTAGAAGGGGCTCTATTCTGCTATATGGCACAGACCTTGGATCTGGTCCAAGTGTTCCT 194

QY 558 GlyAspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyVal 577

Db 193 GGTGATCTCTGGACGACAGGCTAAACCTGGAGTGGATATGATGGTGGTGTGTTGTTATT 134

QY 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys 597

Db 133 CGCGTAGACTCGCCCTCGGACCTTTAAGATTAGACTATGCCTTC--AATGACCAGCGG 77

QY 598 LysIleArgTrpHisIleSerLeuGlyThrArg 608

Db 76 ACCGGAGGCTTTCACCTTTCGGGTTCGCCCTAAGG 44

RESULT 2

AV557254/c 607 bp mRNA linear EST 06-SEP-2000

LOCUS AV557254 Arabidopsis thaliana green siliques Columbia Arabidopsis

DEFINITION thaliana cDNA clone SQ064a09F 3', mRNA sequence.

ACCESSION AV557254

VERSION AV557254.1 GI:8728669

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 607)

AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)

MEDLINE 20363093

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

Location/Qualifiers

1..607

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="SQ064a09F"

/clone_lib="Arabidopsis thaliana green siliques Columbia"

/tissue_type="green siliques"

/note="Vector: phuescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 192 a 172 c 113 g 130 t

ORIGIN

Alignment Scores:

Pred. No.: 6.82e-05 Length: 607

Score: 129.00 Matches: 49

Percent Similarity: 40.12% Conservative: 16

Best Local Similarity: 30.25% Mismatches: 55

Query Match: 4.07% Indels: 42

DB: 10 Gaps: 8

US-09-857-669-2 (1-609) x AV557254 (1-607)


```

Oy 558 -----GlyAsPalaAlaalaAsnPhelYsArgMetLysLeuLysHisGlySer 573
Db 247 CCTAAAGTTCTCGTGGTCCAGCCGAGCTCGTGAAGCCAGGAGTTCGGCTAT 188
Oy 574 GlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGly 593
Db 187 GGTGTGGCGTCCGCGTGGACTCCCTCTGGACCTCTACGCGCTTGAG-----TACGCC 134
Oy 594 HisSerAspLysLysIle-----ArgTrpHisIleSerLeuGlyThrArg 608
Db 133 TTATGATAGCAGCAGCAGCAGATTCACTTGGTGTGGCTACAGA 86
RESULT 6
LOCUS AK002711
DEFINITION Mus musculus adult male kidney cDNA, RIKEN full-length enriched
sequence.
ACCESSION AK002711
VERSION AK002711.1 GI:12832895
KEYWORDS HTG; CAP trapper.
SOURCE Mus musculus
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:0610030001.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Muramatsu,M., and Hayashizaki,Y.
Normalizati and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Muramatsu,M., and Hayashizaki,Y.
Normalizati and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Aizawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yananaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaudo,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Momberts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.

```

and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5 (bases 1 to 2776)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yananaka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGCGCGCCCACTGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0. Second strand cDNA
was prepared with the primer adapter of sequences[5'
GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.
Location/Qualifiers
1..2776
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:0610030001"
/db_xref="MGD:MG1:1907168"
/db_xref="taxon:10090"
/clone="0610030001"
/sex="male"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..2776
/gene="pfkm"
52..2294
/gene="pfkm"
/note="data source:MGD, source key:MG1:97548,
evidence:ISS
phosphofructokinase, muscle
putative"
/codon_start=1
/protein_id="BAB22303.1"
/db_xref="GI:12832896"
/db_xref="MGD:MG1:97548"
/translation="MTHEHHAAKTLGKIAVLTSQDQAGMNAAVRVVVGIFT
GARVFFVEHYGVLVDGGEHREATWESVMMQLQGVIGSARCKDFREGRRAA
HLNVRKGTNLNLCYIGGDSLGSADTFRSEWDLNLDQDKGITAEATKSYNLING
LVLSIDNDFCGDTMGTDLSAHLRIEIVDAITTTTQSHQRTFVLEWGRGKYLALV
TSLSCGADWVFIECPDDDWFEHLICRLSETRTGRSLNIIIVAEIGADKNGKPIITS
EDIKNLVVKRLGIDRVTVLGVQRGTPSPAFDRITLGRMGVEAVMALLEGTPDTPAC

D	b		533	C T C T C G T A G G G A C C C A G C A G G G C C A T C C A G G C C G A C G G C C A C G - - - C G G C T C	599
Q	y		213	A s n V a l v a l A s p S e r G l y P r o f i l e A l a - - - P h e G l y A s p P h e G l u I l e T h r G l y	231
D	b		590	G C C G T G C T G A C A G A C G C A C C E C C C A T C G C G C T C T C A C G C A C A T T G A G A T C - - - - -	643
Q	y		232	T h r G l n A r g T y r P r o G l u G l n l e v S e r G l y L e u A l a A r g - - - - - P h e G l n	247
D	b		644	- - - - - T A T A A G N A C G A C A T N N N N N A C A A C T G C A	694
Q	y		248	P r o G l y T h r P r o T y r - - - - -	252
D	b		695	C C T G G A T T A C C T T A T G T T G A G G A G C A A T T A C C A A T G C T G T G A C T G G T T O A T T G T G G G	754
Q	y		253	A s p L e u A s p L e u L e u - - - - - L e u A s p P h e G l n G l n A l a L e u G l u G l n A s n G l y H i s T y r	270
D	b		755	G A C T T G G A G G T T A T A C A C C A A T C A A G T A C A T G A T G C T C T A G A T C A G A T - - - C G C C T G	811
Q	y		271	S e r G l y A l a S e r V a l G l n A l a A s p P h e A s p A r g L e u G l n G l y A s p A r g V a l - - - - -	287
D	b		812	T C T C C A G C A C A G C T G C G T A A G A G T T T G C C A G G C G C A A T G C T G A T G A C A G T A T T T G C C T T	871
Q	y		288	- - - - - P r o V a l - - - - - L y s V a l S e r V a l G l u V a l L y s A r g H i s	299
D	b		872	C A G C T T C C G C A A T C C A G T A C A C A A T G G C A T G C T T C T T A T G C C G A C A C A C A A A C G T	931
Q	y		300	L y s L e u C l u t h r G l y I l e A r g L e u A s p - - - - - S e r G l u T y r G l y L e u G l y G l y L y s I l e	317
D	b		932	C T C C T T G A G A T G G T T A T N N N N C C T G T T C T T C T C C A T C C A C T G G G A G G A - - - - -	985
Q	y		318	A l a T y r A s p T y r T y r A s n L e u P h e A s n L y s G l y T y r I l e G l y S e r V a l T y r A s p M e t	337
D	b		986	- - - - - T T C A A A A G C A G A T G A T G C C T T A G T T G C A G A A T G	1024
Q	y		338	A s p L y s T y r C l u t h r T h r L e u A l a A d A g l y I l e S e r G l n P r o A r g A s n T y r A r g G l y A s n	357
D	b		1025	A A G C A A C A T G A A A G G T T C T T G A G A A G T G C T C A A C C C A G A A T C A A C T G T T G T G C G	1084
Q	y		358	T y r T r i p t h r S e r A s n V a l S e r T y r A s n A r g S e r T h r T h r G l n A s n L e u G l u L y s A r g A l a	377
D	b		1085	A T C T T T C C C T C T C C A A T G C A T T A T G C T G G C C A A C T A G G T G C A G - - - - -	1129
Q	y		378	P h e S e r G l y G l y I l e T r p T y r V a l A r g A s p A r g A l a G l y I l e A s p A l a A r g L e u G l y A l a	397
D	b		1130	- - - - - T G C A T G C T A A G G C T C G T - - - - - A T T A A T G C T - - - - - G G T G C A	1162
Q	y		398	G l u P h e L e u A l a G l u G l y A r g L y s I l e P r o G l y S e r A s p I l e A s p L e u G l y A s n S e r H i s	417
D	b		1163	A A T T T C T A T A T T C T T G G A G G G A T - - - - - C C T G C T G G T A T G A C C A T	1204
Q	y		418	A l a T h r - - - - - M e t L e u T h r A l a S e r T r p L y s A r g G l n L e u L e u A s n A s n V a l	433
D	b		1205	C C C A C G G A A A A G G C A C T C T A T G A T G C T G A C C A C C G G A A G A G G T T T G A C C A T G C T	1264
Q	y		434	L e u H i s P r o G l u A s n G l y H i s T y r L e u A s p G l y L y s I l e G l y T h r T h r L e u G l y T h r P h e	453
D	b		1265	C C T G C C T C G A G A G G C T C A A C A C T C C T T C C T T T A A G G T G - - - - -	1303
Q	y		454	L e u S e r S e r T h r A l a L e u I l e A r g T h r S e r A l a A r g A l a G l y T y r P h e P h e T h r P r o G l u	473
D	b		1304	- - - - - G C T C A T A T G A C A A A A A A A A A A A A A A A T G A T T C T T T G A T C C A T C A	1351
Q	y		474	A s n L y s L y s L e u G l y T h r P h e I l e L e A r g G l y G l n A l a G l y T y r T h r V a l A l a A r g A s p	493
D	b		1352	A G G A A A C A T A T T C C T C T T C - - - - - A T C T C T G C A C A A A G A T C C G C A C T T C T T G C C A A G A A C	1408
Q	y		494	A s n A l a A s p V a l P r o S e r G l y L e u M e t P h e A r g S e r G l y G l y A l a S e r S e r V a l A r g G l y	513
D	b		1409	C G A G A G A G T N N N N A G A T G T T T A T G T C C C C G G T G C T G C A A G G T A C T C G T T G A A T A C	1468
Q	y		514	T y r G l u	515
D	b		1469	T A T G A	1474

```

RESULT 8
AW053211/c
LOCUS      588 bp      mRNA      linear      EST 20-SEP-1999
DEFINITION 614073B08.xl 614 - root cDNA library from Walbot Lab Zea mays cDNA,
mRNA sequence.
ACCESSION AW053211
VERSION   AW053211.1  GI:5915570
KEYWORDS  EST.
SOURCE    Zea mays.
ORGANISM  Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 588)
AUTHORS    Walbot V.
TITLE      Zea mays ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
COMMENT    Unpublished (1999)
CONTACT    Walbot V
DEPARTMENT Department of Biological Sciences
STANFORD UNIVERSITY
855 California Ave, Palo Alto, CA 94304, USA
TEL: 650 723 2227
FAX: 650 725 8221
EMAIL: walbot@stanford.edu
PLATE: 614073 row: B column: 08.

FEATURES             source
     location/Qualifiers
     1..588
     /organism="Zea mays"
     /cultivar="W23"
     /db_xref="taxon:4577"
     /clone_lib="614 - root cDNA library from Walbot Lab"
     /tissue_type="root"
     /dev_stage="3-4 days old"
     /lab_host="XLOUR"
     /note="Organ: root; Vector: pBluescriptII SK+; Site 1:
     ECORI; Site 2: XhoI; 3-4 days old root tissue from Walbot
     Lab (LM)"
BASE COUNT  182 a 156 c 112 g 138 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0103      Length:      588
Score:          111.00      Matches:     36
Percent Similarity: 42.22%      Conservative: 21
Best Local Similarity: 26.67%      Mismatches:  54
Query Match:     3.50%      Indels:     24
DB:              10      Gaps:       5

US-09-857-669-2 (1-609) x AW053211 (1-588)
QY  476 LysLeuGly---ThrPheIlelleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsn 494
DB  573 AAGTTGGGTCGACATTTTGGTTACTAGTTTACAGGGCGGATCCATTGTTGGGACATG 514
QY  495 AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyr 514
DB  513 GCACCTTATCAAGCT-----TTTGGCATAGGTGGAGCTTGGTACTGCCAGGCTAT 463
QY  515 GluLeuAspSerIleGlyLeuAlaGlyProAnGlySerValLeuProGluArgAlaLeu 534
DB  462 -----GGTGAGGGCGCTGTTGGTTCAGGAAGACTATCT 430
QY  535 LeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPhe 554
DB  429 CTAATTGCTAAGTCGAATACGATTCCCTTGGTGAAGCATCTCGAGGGTCTATTTC 370
QY  555 HisAspMetGlyAspAlaAlaAlaAsnPheLysArgMet-----LysLeuLys 570
DB  369 ATGAGCTGTGGGAGTCACTAGGCTCTGCTGCTCATGTCCTGGTAAATCCAGCTCTCGCG 310
QY  571 HisGlySer-----GlyLeuGlyValArgTrpPheSerProLeu 583

```

```

DB  309 CAGGGAAACAGGATTGGTATTGGATTGGATACGAGCTTCACATTCAACACTGACCTA 250
QY  584 AlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys 598
DB  249 GGTTCAGATCCGTGTCGACTACGCTATGAATGCTATTTAACCAGCAA 205

RESULT 9
AW108685/c
LOCUS      503 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays PCO127797 mRNA sequence.
ACCESSION AW108685
VERSION   AW108685.1  GI:21211868
KEYWORDS  HTC.
SOURCE    Zea mays.
ORGANISM  Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 503)
AUTHORS    Hainey C.F., Dolan M., Miao G.H., Vogel J.M., Whitsitt M.S.,
Arthur L.W., Hanafey M., Morgante M. and Tingey S.V.
TITLE      Zea Mapping Project/DuPont Consensus Sequences
JOURNAL    Overgo Probes
REFERENCE  2 (bases 1 to 503)
AUTHORS    Coe E.C.
TITLE      Direct Submission
JOURNAL    Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES             source
     location/Qualifiers
     1..503
     /organism="Zea mays"
     /db_xref="MaizeDB:637445"
     /db_xref="taxon:4577"
     /clone_lib="PCO127797"
     /clone_lib="Maize Mapping Project/DuPont Consensus
     Library"
     /note="this sequence is part of a project of EST
     assemblies resulting from the application of public
     contigs to seed DuPont contigs; this resource was
     assembled by DuPont as part of a collaboration for the
     overgo addressing of BACs in conjunction with the Maize
     Mapping Project"
BASE COUNT  151 a 137 c 99 g 116 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0124      Length:      503
Score:          109.50      Matches:     38
Percent Similarity: 42.24%      Conservative: 11
Best Local Similarity: 32.76%      Mismatches:  43
Query Match:     3.45%      Indels:     24
DB:              11      Gaps:       4

US-09-857-669-2 (1-609) x AW108685 (1-503)
QY  506 GlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn 525
DB  496 GTTGGACAAATAGTGAAGAGGATATGAA-----GAA 464
QY  526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPhe 545
DB  463 GGTGCTGTGGCTCTGGCGCTTCTTATGCTGTAGGTAGTGGTGAAGTGTCTATGCCCATG 404
QY  546 ThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
DB  403 TTTGGCCCACTGCAAGCGCTGTTTGGAGACATCTAGTGTGCTTGTGTCCT 344
QY  558 -----GlyAspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGly 574
DB  343 AAGTTCTCTGGTGACCCAGCCGAGCTCGTGGAAACCCAGGAGTGGCTTGTGCTATGGT 284

```

Qy	575	LeuGlyValArg-TrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHi	594
Db	283	GTTCGGCTCGCGTACCTCCCTCTGGAGCCTCTACCGCTGAG-----TACGCGCT	230
Qy	594	sSerAspLysLysIle---ArgTrpHisIleSerLeuGlyThrArg	608
Db	229	TAATGATAGCAAGCAAGCAGATTTCATCTTGGTGTTCGTGGCTACAGA	184
RESULT	10		
LOCUS	AK004789		
DEFINITION		3045 bp mRNA linear HTC 19-JAN-2002	
VERSION			
KEYWORDS			
SOURCE		Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200015H18; transformed mouse 3T3 cell double minute 1, full insert sequence.	
ACCESSION	AK004789		
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

LPCRAPPYCHPSSRIQGRRLRDPFQHNMGKPRPTNNLQLHPHDAFNDEBDAURLSEISGA
 SAVSSURAFQTLARAQKRKENFNGKP*
 polyA_signal 3027..3032
 /gene="Mdm1"
 /note="putative"
 polyA_site 3045
 /gene="Mdm1"
 /note="putative"
 BASE COUNT 877 a 689 c 747 g 732 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.302 Length: 3045
 Score: 108.00 Matches: 120
 Percent Similarity: 35.52% Conservative: 64
 Best Local Similarity: 23.17% Mismatches: 200
 Query Match: 3.40% Indels: 136
 DB: 11 Gaps: 29
 US-09-857-669-2 (1-609) x AK004789 (1-3045)
 Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIle---LeuSerAspGlyAsnLeuAla 152
 Db 52 GTCGGCGGGCAGCGCTCCGCAGCCAGCGGCAGCATCGCGTGCCTTCAAGGGCTGAGT 111
 Qy 153 GluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGln-ProVal----- 166
 Db 112 GAATACCAGAGAACTTCCTG-----TGGAAAAAGCTTATTTGTCCAGAGCTTTATAAT 165
 Qy 167 -----GlySerAspPheAspGlnAspSerTrpGluAsnSe 178
 Db 166 CCTCATGTTGGACAAAAGTACTCATGGCAGGACTTAGATCGGATCGATCTGGGATCAGC 225
 Qy 178 rLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeu----- 196
 Db 226 AAGAACCAGGTTTA-----TTTCAAAAAGAGAGATTCCCTACCATGACCTCAGATT 279
 Qy 197 -GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVala 216
 Db 280 TCAAAATACCTGGAGTGGACCGAACCGCTCAGAAAGAGGATACGCTTGTC-----CA 333
 Qy 216 lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPr 236
 Db 334 CCAGAACCCAGGGCTTGG-----GAAGCCCAAGCCACAGAGGGCTGAG 378
 Qy 236 oGluGlnIleValSerGlyLeuAlaArgPheGln-----ProGlyThrProTy 252
 Db 379 CAAGGAGAAGATGCCAATCAAGAGCAGATTCTCTCACTAGAGGCTCCAGGGTTCCCAAG 438
 Qy 252 rAspLeuAspLeuLeuLeu-----AspPheGlnGlnAlaLeuGlu-- 265
 Db 439 AGAACTCGGTCTCATCTTCGGGAGCTCGAGAGCTGAAGGGGTTCAGAC- ACTGTGGA 497
 Qy 266 ---GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnG 284
 Db 498 GCACCAGGGTGTCCAGGAGAACCCATGCCCGCAGTTAGCCGGATGTGGAGCTGACACCTTC 557
 Qy 284 yAspArgValProValLysValSerVal---ThrGluValLysArgHisLysLeuGluTh 303
 Db 558 CAGCAAAACCTCTCTCCAGCATAGATCCCGGATGTGGATAGGCAT----- 606
 Qy 303 rGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAs 323
 Db 607 -----CTTCGTAGAAAGCTGGATTGGCC-----GTTGTCCCAAGAAATAATGC 650
 Qy 323 nLeuPheAsnLysGlyTyrIleGlySerValIleTrpAspMetAspLysTyrGluThrTh 343
 Db 651 CTTGAGAAATCTCAATACCAAGCAGCTTGTGTGGAGACTTCTAAAGAAACGGCTCC 710
 Qy 343 rLeuAlaAlaGlyIleSerGlnProArgAsn-----TyrArgG 356
 Db 711 AGTGTTTCATCCCAATCAGGTTTTCGCTTAATAAAGCCAAATTAATTCACAGTTC 770

Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Tissue Procurement: Dr. Igor Dawid
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.lni.gov
 Plate: L1AM10804 row: 1 column: 19
 High quality sequence stop: 605.
 Location/Qualifiers
 1. 928
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:488026"
 /clone_lib="NICHD XGC Emb2"
 /tissue_type="embryo, stage 17/19"
 /lab_host="DH10B (phage-resistant)"
 /note="vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 2.1 kb. Constructed by Life Technologies."
 BASE COUNT 256 a 207 c 243 g 222 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.172 Length: 928
 Score: 103.50 Matches: 72
 Percent Similarity: 36.48% Conservative: 40
 Best Local Similarity: 23.45% Mismatches: 89
 Query Match: 3.26% Indels: 106
 DB: 14 Gaps: 17

US-09-857-669-2 (1-609) x BQ723872 (1-928)

Qy 98 AsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGlu 117
 Db 34 AACATCTTGTCCGTCATGAGGAGCCAGGG----- 63
 Qy 118 LysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137
 Db 64 -----ATCGAGCCAGGCGCCAGAGACTTACGTGGCTCTTTTA 99
 Qy 138 ---GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArg 156
 Db 100 ACCGCAATTTGCCGAGAGAGGGGATATT-----AACAAATATCAAGAG----- 141
 Qy 157 AsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGlu 176
 Db 142 ---ACCTTAGAAGACGTGGAAAGAAATCAGGAAGCCCTCAGACCGAGATTGATGCAA 198
 Qy 177 AsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro----- 192
 Db 199 -----GTCATCTGGAGTCTGGCTAAGGCTGGATACCCCTCAGTATGTTCCAG 243
 Qy 193 -----LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
 Db 244 GACATCATGGCGCGCATGATATGACCGAGATACATA---CCAGATGCGGATGAACCTC 300
 Qy 211 AspLeuAsnValValAspSerGlyArgProIleAlaPhe----- 224
 Db 301 TCTTTGAGTCTGTACACACAGGTTATGAGATGTAGCTTCTAGTGTGAATCCCTTC 360
 Qy 225 -----GlyAspPheGluIle 229
 Db 361 TCGGCTGCCTCGTACAGATTCGAGACGCGGATCTCTACAGCAGCGGAACCTCTCTCTG 420
 Qy 230 ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGly 249
 Db 421 CGTCACTCGCTCAGTTTG---GAGCAGCCAGCAAGTAAGGTCAAGCAGTTCTGTGTGGT 477

Qy 250 -----ThrProTyrAspLeu----- 254
 Db 478 TTGAAGAAGCAATTTTGCATTCGGCACCTCTGCAGTTTGTCTGTACTGTCTCTCGAT 537
 Qy 255 -----AspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGly----- 268
 Db 538 GCCAAAAAACAGATTTAGCTCTCATGAATGTCTGAAGAGAGAGAGATTCCT 597
 Qy 269 -----HisTyr-SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAs 285
 Db 598 GTCAGGCCACACTACTGCTGCCTCTCTGATCAGCTTCC---AAAAAGAAAAAATTCA 654
 Qy 285 pArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyI 305
 Db 655 GAGGTACCCATTAAGTAATT-----AAAGCGCTGTCTGAAATGGGAGT 699
 Qy 305 eArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPh 325
 Db 700 GGAGCTAGACGTTGAG-----ACCTACTCCCACTATGCGCTG-- 736
 Qy 325 eAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAl 345
 Db 737 -----ACTGTGTTAATGATGATACAGAGCGGCTCGTCACACA---GC 774
 Qy 345 aAlaGlyTyrSerGlnPro 351
 Db 775 TGCAGGAGAAAGCGTGGCC 793
 RESULT 15
 AI363592/c
 LOCUS
 DEFINITION SWOVL2CAS01E04SK Onchocerca volvulus L2 larvae cDNA (SAW98MLW-OvL2)
 Onchocerca volvulus cDNA clone SWOVL2CAS01E04 5', mRNA sequence.
 ACCESSION AI363592
 VERSION AI363592.1 GI:4115213
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Onchocerca volvulus.
 Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 REFERENCE
 1 (bases 1 to 423)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Genes expressed in L2 larvae of Onchocerca volvulus
 Unpublished (1999)
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: phluescript SK.
 Location/Qualifiers
 1. 423
 /organism="Onchocerca volvulus"
 /db_xref="taxon:6282"
 /clone="SWOVL2CAS01E04"
 /clone_lib="Onchocerca volvulus L2 larvae cDNA
 (SAW98MLW-OvL2)"
 /dev_stage="L2"
 /lab_host="XLI-Blue MRF"
 /note="vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. mRNA was
 prepared from approximately 9,000 L2s isolated from
 infected mosquitoes from Kumba, Cameroon and converted to
 double-stranded cDNA using reverse transcriptase and
 oligo(dT) followed by RNase H and DNA pol I. The library
 has 7.3 x 10E4 independent recombinants and the average
 insert size is approximately 1kb. The library was
 constructed by Michelle Lizotte-Waniewski. The library is
 available from Dr.S.A.Williams, email: genome@smith.edu."
 BASE COUNT 99 a 119 c 111 g 90 t 4 others

gene /dev_stage="10 days embryo"
 1. .2936
 /gene="Mdm1"
 114. .2240
 CDS /gene="Mdm1"
 /note="data source:MGD, source key:MGI:96951,
 evidence:ISS
 putative
 transformed mouse 3T3 cell double minute 1"
 /codon_start=1
 /protein_id="BAB27830.1"
 /db_xref="GI:12848104"
 /translation="MPVREKGLSEYORNEFLMKKSYLSESNPSVGOKYSWAGLKSDQL
 GITKEPGFISLRVPKTRSHSADSRAGVSDTVKEHQGVTRSHAPVADVLRLPSK
 ANDEVALSASRPKTRSHSADSRAGVSDTVKEHQGVTRSHAPVADVLRLPSK
 QPLSQSLDPLRHLRKKAGLAVPTNNALNSEYQRFVWKTSKESAPVAFNQVFR
 NKSOIQPOGFTTETETKRNFKGLTPVKEPKSREYLGKSSLEMLTPVKRADDPL
 DLEVDMASEDSQSVKPKASMRHORLGVNSEYRAKFLSPAQYFYKAGAWTRVKNLS
 NOVKELEKAESEYKRVQGFHFGRDLNQLTMSDSNCCWDSVTSSEGTVSSNIRALD
 LAGDLTNHRTPOKHPPTKLEERKVASGEQPLKNSRLEMPPEAPASVRRKLANDAES
 TKEDTOPEAREEDGEREGODKOTCAVELEKPDOTPKADRLTEGESSETSSVSGGG
 RLTPRLRELGIORTHDLTTPAVGGAVLSPSKVPGLLEQRRRRASSQGLLETKAD
 ITRKGRPMSLTSPAAAGMKTVDPILPREDCEANVLRFDATLPVSKILDRQSTPGQ
 LPPCAPPYCHPSRIQGRLDPEFOHNMGRKPTRNNLQLHPDAFNDDEADRLSEISAR
 SAVSSIRAFOTLARAQKRKENFWGKP"
 2921. .2926
 /gene="Mdm1"
 /note="putative"
 2936
 /gene="Mdm1"
 /note="putative"
 2936
 /gene="Mdm1"
 /note="putative"
 2936

polyA_signal

polyA_site

BASE COUNT 832 a 673 c 740 g 691 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.17 Length: 2936
 Score: 103.00 Matches: 102
 Percent Similarity: 34.51% Conservative: 55
 Best Local Similarity: 22.42% Mismatches: 176
 Query Match: 3.25% Indels: 123
 DB: 11 Gaps: 25

US-09-857-669-2 (1-609) x AK011769 (1-2936)

Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIle---LeuSerAspGlyAsnLeuAla 152
 Db 81 GTCGCGGGGAGGCTCCGACGCGGAGGCGGACATGCGCGGTCTCAAGGGGCTGAGT 140
 Qy 153 GluTyrTrpArgAsnAlaLeuGluAsnTrpGlnGln-ProVal----- 166
 Db 141 GAATACACAGAACTTCCTG-----TGGAAAAGTCTCTATTTGTCAGAGTCTTATAAT 194
 Qy 167 -----GlySerAspPheAspGlnAspSerTrpGluAsnSe 178
 Db 195 CCCTCAGTGGGACAAAAGTACTCATGGGCGAGACTTAAATCGGATCAGTTGGGGATCAGC 254
 Qy 178 rlyThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeu----- 196
 Db 255 AAAGAACACAGGTTTFA-----TTTCAAAAAGAGAGATTCCCTACCATGAGCCCTCAGATT 308
 Qy 197 -GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValIva 216
 Db 309 TCMAATACCTGGAGTGAAGGAAACCTCAGAAAGAGGATACGCTTGTC-----CA 362
 Qy 216 lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPr 236
 Db 363 CCAGAACCCCGGCTTGT-----GAAGCCCAAGCCACAGAGGCTCAG 407
 Qy 236 oGluGlnIleValSerGlyLeuAlaArgPheGln-----ProGlyThrProTy 252
 Db 408 CAAGGAGAAGATGCCAATCAAGAGCAGTTCTCTCCTACCTAGAGGCGCTCCAGGCTTCCCAAG 467

Qy 252 rAspLeuAspLeuLeuLeu-----AspPheGlnGlnAlaLeuLeuGlu-- 265
 Db 468 AGAACTCGGTCTCATTTGCGGACTCGAGAGCTCAAGGGGTTTCAGAC-ACTGTGGA 526
 Qy 266 ----GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnG 284
 Db 527 GCACCAAGGGTGTACAGGAAGCCATGCGCCAGTTAGCGCGATGTGAGACCTTC 586
 Qy 284 yAspArgValProValLysValSerVal---ThrGluValLysArgHisLysLeuGluTh 303
 Db 587 CAGCAACAACCTCTCTCCAGACATAGATCCAGGTGGATAGGCAT----- 635
 Qy 303 rGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyAs 323
 Db 636 -----CTTCGTAAGAAGCTGGATTGGCC-----GTTGTCTCCACGAATAATGC 679
 Qy 323 nLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThTh 343
 Db 680 CTTGAGAAATTTCTGAATACCAAGGCAGTTTGTGGAGAGCTTCTTAAGAAAGCGCTCC 739
 Qy 343 rLeuAlaAlaGlyIleSerGlnProArgAsn-----TyrArgG 356
 Db 740 AGTGTTCATCCCAATCAGGTTTCCGTAATAAAGCCAAATTTATCCACAGTTC 799
 Qy 356 yAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysAr 376
 Db 800 CAATACATTCACCCAGGAGCTGAATAC-----AAGCG 832
 Qy 376 gAlaPheSerGly-----GlyIleTyrTrpValArgAspArgAlaGlyIleAspAlaAr 394
 Db 833 AAATTTCAAGGTTTAACTCCAGTGAAGGAACCAAGTCAAGAGAGATATTTGAAGGAAA 892
 Qy 394 gLeuGlyAlaGluPheLeuAlaGluGlyArgLysIlePro-----GlySerAspI 411
 Db 893 CAGCAGTCTGAGATGCTGACTCCAGTAAAGAGGACAGATGAGCCTTTAGACTTAGAAGT 952
 Qy 411 eAspLeuGly-----AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgG 428
 Db 953 AGACATGCGGTCCGGAAGACTCAGACCTGCTGTAAGAAGCCTGCTTCATGAGACACCA 1012
 Qy 428 nLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyTh 448
 Db 1013 AAGCCTTGAAAGTGT-----AATCTCAATAT----- 1040
 Qy 448 rThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTy 468
 Db 1041 ----AGACAAAGTTCTCTGAGCCCA-----GCCAGTA 1069
 Qy 468 rPhePhe-----ThrProGluAsn-----LysLys 476
 Db 1070 TTTCTATAAGCTGGAGTTTGGACCCGGGTGAGAGAACCTGTCAACACGAGTTAAGA 1129
 Qy 476 sLeu-----GlyThrPheIleIleArgG 484
 Db 1130 GCTCCGAGAAAAGCCGAATCTTACAGGAAGCGAGTTTCAGGGGACACATTTTCTCGGA 1189
 Qy 484 yGlnAlaGlyTyrThrValAlaArgAspAsnAla-----AspValProSerGlyLeuMe 502
 Db 1190 CCATCTGAACCAAGATTATGTGCGGACAGCAACTGCTGTTGGGACGTCCTCTCAGTCA 1249
 Qy 502 tPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
 Db 1250 CTCGGAAGCCCGTCAGTAGACATCCGACACTGGATCTTT 1292

RESULT 17

BM438066

LOCUS

DEFINITION

BM438066

ACCESSION

VERSION

KEYWORDS:

BM438066 451 bp mRNA linear EST 31-JAN-2002
 VWA030A02.55627 An expressed sequence tag database for abiotic
 stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera
 cDNA clone VWA030A02 5, mRNA sequence.
 BM438066
 BM438066.1 GI:18459788
 EST.

SOURCE Vitis vinifera.
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.

REFERENCE 1 (bases 1 to 451)
AUTHORS Cramer,G.R. and Cushman,J.C.
TITLE An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay
JOURNAL Unpublished (2002)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 030 row: A column: 02
Seq primer: T3 20mer
High quality sequence stop: 451.

FEATURES
source
1..451
/organism="Vitis vinifera"
/db_xref="taxon:29760"
/clone="VVA030A02"
/clone_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"
/tissue_type="leaf"
/dev_stage="juvenile and adult"
/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda Uni-zapXR vector and cDNA synthesis kit."

BASE COUNT 92 a 89 c 120 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0998 Length: 451
Score: 101.50 Matches: 35
Percent Similarity: 41.44% Conservative: 11
Best Local Similarity: 31.53% Mismatches: 36
Query Match: 3.20% Indels: 29
DB: 13 Gaps: 5

US-09-857-669-2 (1-609) x BM438066 (1-451)

QY 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArg 532
|||||||
Db 1 GGTATGAA-----GAAGTGCTGTGGGCTCAGTCCG 33

QY 533 AlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAla 552
|||||||
Db 34 TCTCATGTGGTGGTAGTGAGAAAATTTCTTCCCTTGTACGGCCCAATTAGGAGGGCT 93

QY 553 ValPheHisAspMet-----GlyAspAlaAla 561
|||||||
Db 94 CTTTGTGCTGATGTGGAACCATCTTGGATCAGGCCCACTGTGCTGGCCATCTCTGCA 153

QY 562 AlaAsnPheLysArgMetIysLeuLysHisGlySer-----GlyLeuGlyVal 577
|||||||
Db 154 GGG-----GCCAGGCTAAAGCCTGGAAGTGGATATGGATATGGAAATT 201

QY 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys 597
|||||||
Db 202 CCACGTGGACTTCCTTTAGGGCTCTGGCACTTGAATATGCATTTAATGACCAGCAGGCA 261

QY 598 LysIleArgTrpHisIleSerLeuGlyThrArg 608
|||||||
Db 262 CAG---AGGTTTCACTTTGGGTTGGTCAACCA 291

RESULT 18
BQ876247
LOCUS BQ876247 1031 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT_8742755 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6391121
5', mRNA sequence.
ACCESSION BQ876247
VERSION BQ876247.1 GI:22268253
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1031)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13878 row: n column: 18
High quality sequence start: 103
High quality sequence stop: 510.

FEATURES
source
1..1031
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6391121"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

BASE COUNT 249 a 240 c 294 g 246 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.356 Length: 1031
Score: 101.50 Matches: 93
Percent Similarity: 32.76% Conservative: 40
Best Local Similarity: 22.91% Mismatches: 129
Query Match: 3.20% Indels: 145
DB: 14 Gaps: 22

US-09-857-669-2 (1-609) x BQ876247 (1-1031)

QY 125 HisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAsp 144
|||||
Db 31 CACAGCTCTCCCT---CCAAAAAAGAGCTGGAGCGGC-----CGGAAT 75

QY 145 IleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGln 164
|||||
Db 76 ATTCGAACGAGCATACCCCTCAGGAATTTAC-----CACTACTATCAGCTT 123

QY 165 ProVal-----GlySerAspPheAsp 171
|||||
Db 124 CCAGTCTGCTGCTCCTGAGAGATCCGCCAAGAGCTTAGTTGGGTGAGCTTCTGGAT 183

QY 172 GlnAsp-SerTrpGluAsnSerLysThr---SerValLeuGlyAlaValThrArgLysG 190
|||||
Db 184 GGGGACCGGAATGGCGGAATCTTTGTACGAGATCCGCTTTCGGGAGAAATGCGAGAAGA 243

QY 190 yTyrProLeu-----AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 207
|||||
Db 244 ATTCTGTGTACATGCTCAGCTCAGTCTGTGCACAGGTGGAGCAGCTGCGCCAGGCCATTGAG 303

```

Oy 207 rAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPh 227
Db 304 GAATGTATTACTTTGCGTATTTT-----GATTT 338

Oy 227 eGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaAArgPheG1 247
Db 339 GCCAATCGTGGTTTGGCGCTACATGGAAG-----AGTGGC-----TTCTT 383

Oy 247 nProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAs 267
Db 384 GCCA-----CA 389

Oy 267 nGlyHisTyrSerGlyAlaSerValGlnAlaAspPheA-----ArgLeuGlnGlyAspAr 286
Db 390 TAGCCACAAGATAGCCCTCTGGACACATTTGGACTTCCACCTAGATATTCATGGAGATCG 449

Oy 286 gValPro---VallysValSerValThrGluValLysArgHisLysLeuGluThrGlyL1 305
Db 450 AATTATATTGCAATGTTTCGGTGGGAGCTCAAGCCCAAGCTTGGAT---GGGCT 506

Oy 305 eArgLeuAspSerGluTyrGlyLeu----- 313
Db 507 GCGGTCTGATCAATGTCTGGGCTTACTCACACTTACGTGTGCGCTGGTCTGNAACAG 566

Oy 314 -----GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysG1 328
Db 567 TCCAGTGGGAGCATCCGCAAGTGCACAAAGCCGCCCAAGGTGAAT-----GG 614

Oy 328 yTyrIleGlySerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyL1 348
Db 615 ATGGGAAGGC-----CTTCTCTTTTTCCT 638

Oy 348 eSerGlnProArgAsnTyrArgGlyAsnTyrTrp-----ThrSerAsnValSerTyrAs 366
Db 639 CCAAAACCCCTCGGAAAAATCCCTTTTGGGGTGGTGGCCCAATTTCCCATTT 698

Oy 366 nArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTyrTyrValAr 386
Db 699 CCAAAACCTTTTCCCATGTTGGGGGGGCGCCCTTTTGGGGGGGTTTTTTTATTAA 758

Oy 386 g-----AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyAr 404
Db 759 CACTTGGGGAGGGGGGGTTTTTTTTTGGGGGGGGCC----- 798

Oy 404 gLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSe 424
Db 799 -----CCCGGGGTAAATTTCCCAAAATTTG----- 828

Oy 424 rTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspG1 444
Db 829 -----CCCGGGGGGTGCCCTTTTCCCAA 854

Oy 444 yLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleAArgThrSerAl 464
Db 855 AAAATATGGGACCCCTTTTGGGACTTTTCCGCGGGGAACC----- 894

Oy 464 aArgAlaGlyTyrPhePheThrPro-----GluAsnLysLysLeuGlyTh 479
Db 895 -----ACACCCCTCTTAAACCCGAAAAAACCAGCGGAGC 935

Oy 479 rPheIleIleArgGly 484
Db 936 GTTCTCTCGGGGGG 951

```

RESULT 19 BH770975/c

LOCUS BH770975 3095 bp DNA linear GSS 01-MAY-2002
DEFINITION LMGtag700 MG1363 Random Sequence Tag Library Lactococcus lactis
subsp. cremoris genomic, DNA sequence.

ACCESSION BH770975

VERSION BH770975.1 GI:20373932

KEYWORDS GSS.

SOURCE Lactococcus lactis subsp. cremoris.

```

ORGANISM Lactococcus lactis subsp. cremoris
REFERENCE Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
AUTHORS 1 (bases 1 to 3095)
TITLE Bolotin,A., Ehrlich,S.D. and Sorokin,A.
JOURNAL Studies of genomes of dairy bacteria Lactococcus lactis
COMMENT Sci. Aliments, (2002) In press
Contact: Sorokin A
Genetique Microbienne
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL1403 is rpoC (92%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 3067.
FEATURES
Location/Qualifiers
1..3095
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"
/note="Vector: pSGM2; Site 1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 915 a 649 c 614 g 912 t
ORIGIN
Alignment Scores:
Pred. No.: 1.93 Length: 3095
Score: 101.50 Matches: 104
Percent Similarity: 34.55% Conservative: 67
Best Local Similarity: 21.01% Mismatches: 187
Query Match: 3.20% Indels: 137
DB: 17 Gaps: 26
US-09-857-669-2 (1-609) x BH770975 (1-3095)
Oy 33 GlyPheAlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysPro---- 51
Db 2583 GGGATTCAAGCTTTTGAACCCAGTCTTAATTGATGCTAAAGCCATCCGTCCTCACCACCTT 2524
Oy 52 -----LysPheProVal 55
Db 2523 GCTTGTGAAGCTATAACGGCGACTTTGACGTGACCAATGGCCATTCATTCGTCATTG 2464
Oy 56 ArgIleAspThrGln---AspSerGluIleLysAspMetValGluGluHisLeu----- 72
Db 2463 TCTGAAGAAGCACAGCTGAAGCAGCTCTTTTGATGCTGTGCTGTAACATATCTTGAC 2404
Oy 73 -----ProLeuIleThrGlnGlnGlnGluValLeuAspLys----- 85
Db 2403 CCTAAAGATGGTAAACACAGTTGTTACACCATCTCAAGATATGCTGCTCTGTAACACTAC 2344
Oy 86 -----GluGlnThrGlyPheLeuAlaGlu-----GluAlaProAspAsn 98
Db 2343 CTTACAATGGGAAGAAAAGGTCGTGAAGCGGAAGGATGATTTTGGCAACTCCTCGAAGAA 2284
Oy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLys 118
Db 2283 GTTGAATCGCAATCGGTAAAC---GGTTAT-----GTACACTTGCATACACGT 2239
Oy 119 AspGlyAlaTyrThrValHisIle-----ThrProGlyProArgThrLysIle 134
Db 2238 ATTGCTATCGGCACAAAATCACTCAATAAACCTTGGACTGAAAAATCAACAGATAAAATC 2179
Oy 135 -----AlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Db 2178 TTGGTAACACACGTGGTAAAGATTATTTCAACTCAATCATTCACAGAGGA-----ATG 2125
Oy 153 GluTyrTyrArgAsnAlaLeuGlu---AsnTrpGlnGlnProValGlySerAspPheAsp 171

```

Db 2124 CCTTACTTGAACGAACCACTGATGTTAACTTGACAACTTCAACTGATGACCGCTTCTTT 2065
Qy 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr----- 167
Db 2064 ATGGATCGGTGCTCAAAACATCAAAAGAGTTTGGCTGGAACTCATCTGTTGCTCCATTC 2005
Qy 188 ArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThr 207
Db 2004 AAAAAAGGATAT-----CTTGGAAATATTATCGCCGAAGTA----- 1969
Qy 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 1968 -----TTCAAAACGTTAC 1957
Qy 228 GluIleThrGlyThrGlnArgTyrProGluGlnIle-----ValSer 241
Db 1956 CGTACACAGCTACATCTGAGTACCTTGACCGTTTGAAGACCTTGGTTACCAACCAATCT 1897
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1896 ACITTGCTGGTTGACTGTTGGGTATCGCT--GATATTCTCTGTTGGAAGATAAACAT 1840
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1839 GAAATCATTCATGCT---GGCATAACCGGTAGAACACATCACTAAACAAATTTCCGTCGC 1783
Qy 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1782 -----GGTTTGATTACTGATGATGACCTTATAATGCC 1750
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321
Db 1749 GTTACTGGTGTGGCTGATGCAAAAGACTCTCGAAAACGATTTGATTGAAGAC--- 1693
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
Db 1692 ---CAAGAAATTGCACAA-----CCCATCGTTATGATGATGAGCTCTGAAGCC 1648
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 1647 CGTGGTAATATTCCCAACTCTCTCAACTGCTGGTATGCTGGATGATGGTGGTCTCT 1588
Qy 362 Asn-----ValSerTyrAsnArgSerThrThrGlnAsnLeu 373
Db 1587 AATGGTAAATCATGCAATTGCCTATCATCTCTAACTTCGTCGAAGCCCTTCTGCTTG 1528
Qy 374 Glu-----LysArg 376
Db 1527 GAAATGTTCTTCAACTCAGGTGCTCGTAAAGGGATGCGCGATAGCGCCCATAGACA 1468
Qy 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIle-----AspAla 393
Db 1467 CGCGACTCAGT-----TACCTTACTCGCTGTTGGTGTGCTGCGCCCAAGATGTT 1417
Qy 394 ArgLeuGlyAlaGluPheLeuAlaGluArgLysIleProGlySerAspIleAspLeu 413
Db 1416 ATCATTCGTCAAAAGGAGCTGCTGACTGACCGTGAACCTTCTGATGCTGATATTGCAACT 1357
Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu----- 429
Db 1356 GGTAAAGAA-----ATGGTTGAACCACTCTCTTGAAGCTTGGTGTGCTGCTTACACT 1306
Qy 430 LeuAsnAsnValLeuHisProGluAsnGlyHisTyrLysArgGly 444
Db 1305 CGTAAATCAGTCTCTCATCCAGAAACTGGTGAATGATTATTGGC 1261
RESULT 20
BQ872838
LOCUS
DEFINITION BQ872838 724 bp mRNA linear EST 15-AUG-2002
OQ116M16.yg abi OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
OQ116M16, mRNA sequence.
ACCESSION BQ872838

VERSION BQ872838.1 GI:22259397
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 724)
AUTHORS Kozik, A., Michelson, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lei, Z.,
Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL http://compgenomics.ucdavis.edu/
COMMENT Unpublished (2002)
CONTACT: Alexander Kozik [R.W.Michelson@vegmail.ucdavis.edu]
Department of Vegetable Crops, R.W.Michelson Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelson@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG116 row: M column: 16.
FEATURES
Location/Qualifiers
1..724
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG116M16"
/clone_lib="OG_ABCDI lettuce salinas"
/lab_host="E.coli"
/note="Vector: pBRCDNASFIAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_lib=OG_ABCDI lettuce salinas
TAG_tissue=flowers pre-fertilized
TAG_SEQ=GCTTGACGGG"

BASE COUNT 255 a 126 c 147 g 196 t
ORIGIN
Alignment Scores:
Pred. No.: 0.315 Length: 724
Score: 100.00 Matches: 46
Percent Similarity: 40.18% Conservative: 42
Best Local Similarity: 21.00% Mismatches: 87
Query Match: 3.15% Indels: 44
DB: 14 Gaps: 9
US-09-857-669-2 (1-609) x BQ872838 (1-724)
Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAACGAGGCACAGCAGCAGGTACATCCATCATCCAAAAG-----GSTACTGAG 53
Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
Db 54 TCACCCAGC-----CAAGCTTTACACACACAAATCT----- 86
Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLys 85
Db 87 -----GTTGTTGTTAGTATTCCAATCATTTGGTGATACAGTCAATCCATCATCCACAA 140
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105

Db 141 CAAACTGAC-----ATGGCAATCAATTGATTGATACCAAGGACTTCTTCTGTCAAGGAG 194
 QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AAAGGTATATTCGGTGTCTGCTACCTTACCTGACAGCAATGTTGGTGCATCAATCATCAA 254
 QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGGTTCTGATTGATTCAGTACCAAGCATTAGAAACAACCAAGTCGGTGTGCGAGGTATC 314
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 315 CCAATTGTTGGGATGACGAGAA-----TCTATTATTACACAAACAGAC 362
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
 Db 363 ATGGCTAGTATCGGTTTGAAGCTCAACTTCGTCGTCGAAGACATGTTTATTAGGA 422
 QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
 Db 423 GCTATTTTATCAGCGGAGCAGCAGGTTGTAGAAGATGTTGGAACACAC----- 470
 QY 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
 Db 471 -----GCAACAATAATCAATTGAAAGAAGTTCTGATTCGTGC 506
 QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
 Db 507 GAAAAAACCTAGAGAAACACATCCAGCATTCGAGATATCCCAATTCAGGATCT 563

RESULT 21

BQ871231 613 bp mRNA linear EST 15-AUG-2002
 LOCUS QG111D13.yg.ab1 OG.ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION QG111D13 mRNA sequence.

ACCESSION

VERSION BQ871231

KEYWORDS

SOURCE EST.

ORGANISM

Lactuca sativa.
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE

AUTHORS

1 (bases 1 to 613)
 Kozik, A., Micheltore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compenomics.ucdavis.edu/
 Unpublished (2002)

JOURNAL

COMMENT

Contact: Alexander Kozik [R.W.Micheltore]
 Department of Vegetable Crops, R.W.Micheltore Lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozike@ucdavis.edu [micheltore@vegmail.ucdavis.edu]

belongs to contig QG_CA.Contig6253, see http://cgdb.ucdavis.edu/
 for details.

Plate: QG11 row: D column: 13.

FEATURES

source

1..613 Location/Qualifiers

/organism="Lactuca sativa"

/cultivar="Salinas"

/db_xref="taxon:4236"

/clone="QG11D13"

/clone_lib="OG.ABCDI lettuce salinas"

/lab_host="E.coli"

/note="Vector: pBRCDNASf1AB; The library was constructed

from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgdb.ucdavis.edu/
 TAG_LIB-OG.ABCDI lettuce salinas
 TAG_TISSUE=flowers pre-fertilized
 TAG_SEQ=GCCTGACGGG*

BASE COUNT 206 a 116 c 129 g 162 t

ORIGIN

Alignment Scores:

Pred. No.: 0.324 Length: 613

Score: 99.00 Matches: 46

Percent Similarity: 40.18% Conservative: 42

Best Local Similarity: 21.00% Mismatches: 87

Query Match: 3.12% Indels: 44

DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ871231 (1-613)

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45

Db 3 GTAACGAGGCACAAAGCAGGTACATCCATCATCCAAAG-----GGTACTGAG 53

QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65

Db 54 TCAGTCAGC-----CAGCTTTAGACACAAATAATCT----- 86

QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85

Db 87 -----GTTGTGTGTAGTATTCCAAATCATGTCATCAGCTCAATCATCCACCAA 140

QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105

Db 141 CAAACTGAC-----ATGGCAATCAATTCATTCATCCACGAGCTTCTGTCAAGGAG 194

QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125

Db 195 AATGGTATATTCGGTGTCTATCATTGACTCAGATGTTGGTGGATCAATCATCCAA 254

QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145

Db 255 ATGGGTTCTGATTGATTCAGTACCAAGCATTAGAAACAACCAAGTCGGTGTGCGAGGTATC 314

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165

Db 315 CCAATTGTTGGGAGCAGCAGAGAA-----TCTATTATTACAAACAACAGAC 362

QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184

Db 363 ATGGCTAGTATCGGTTTGAAGCTCAACTTCGTCGTCGAAGACATGTTTATTAGGA 422

QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203

Db 423 GCTATTTTATCAGCGGAGCAGCAGGTTGTAGAAGATGTTGGAACACAC----- 470

QY 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219

Db 471 -----GCAACAATAATTCATTCGAGATATCCCAATTCAGGATCT 506

QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232

Db 507 GAAAAAACCTAGGAAACACATCCAGCATTCGAGATATCCCAATTCAGGATCT 563

RESULT 22

LOCUS BM560625/c

DEFINITION

5', mRNA sequence.

ACCESSION

BM560625

VERSION

BM560625.1 GI:18805199

BM560625

1171 bp

mRNA

linear

EST 20-FEB-2002

AGENCOURT_6597970 NIH_MGC_98

Homo sapiens

cDNA clone IMAGE:543221

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1905 row: 1 column: 14
 High quality sequence stop: 532.
FEATURES Location/Qualifiers
 source
 1..1171
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:543221"
 /clone_lib="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; site_1: XhoI; site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAGCAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
BASE COUNT 213 a 448 c 325 g 185 t
ORIGIN
 Alignment Scores:
 Pred. No.: 0.876 Length: 1171
 Score: 99.00 Matches: 64
 Percent Similarity: 30.42% Conservative: 16
 Best Local Similarity: 24.33% Mismatches: 64
 Query Match: 3.12% Indels: 119
 DB: 13 Gaps: 13
US-09-857-669-2 (1-609) x BW560625 (1-1171)
 Qy 376 ArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
 Db 1104 CAGTGTCTCACGGGGGGG--TGG-----CGCGGCAATAGATAGCGGTGTG 1060
 Qy 396 GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsn 415
 Db 1059 GCGGTAGATCGGGCGGATCGAGGGGAGAAAATCGCGGGCGT----- 1018
 Qy 416 SerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHis 435
 Db 1018 ----- 1018
 Qy 436 ProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSer 455
 Db 1017 -----GGGTACTGGGCTCGTAAATGGTCCC----- 991
 Qy 456 SerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys 475
 Db 990 -----GGGGTAAGTTATTATTATCCCGCGAGTGCAC 961
 Qy 476 LysLeuGlyThrPheIleleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla 495
 Db 960 TCCGTCGGCAGC-----CGCAGCGGACGCGGCACCTTGTGTGTGTCGGGGCGCGCG 907

496 AspValProSerGlyLeu-MetPhe-----ArgSerGlyGlyAlaSerSe 510
 Db 906 -----TCTGGCATCGCTTTCTTTCAGCGGGGGGATCGAGTGGGGGGGTCCAG 856
 Qy 510 rValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeu-- 529
 Db 855 TATCGCGGTTCACAAATCCACGCGGTGGGG-----GGGCCAAATCGTGTCAATTAAC 802
 Qy 530 -----ProGluArgAlaLeuLeuVal----- 536
 Db 801 ATATGGGTCGGGGGGGGTAAACGGCGGGGGCGCGCTCGCTCTAGTATATCA 742
 Qy 537 -----GlySerLeuGluTyrGlnLeu-----ProPheThrAr 547
 Db 741 GGGGTCACTCGGGGCTCTCGCTTACTACTAGTGGACACCGGGGGTCTATTTTCGGCG 682
 Qy 547 gThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
 Db 681 CGGTTTGGGG-----AGTCCGCGGGCGATTCGATTTCGGGACCGCGCGCTCG 634
 Qy 560 -----Al 560
 Db 633 CGAGTTTGGAGGGCGGAGTGGTGTGGTCCGGGGAAGGCCCGGAGATGCCAAGGCG 574
 Qy 560 aAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPh 580
 Db 573 AGGCAGTCTAGACACGCGTAACGCGATTCGCGCGGATCTCCCCCAGGATCTTGTGTGGCG 514
 Qy 580 eSerPro 582
 Db 513 ATCCCA 507
RESULT 23
BO875918
LOCUS BO875918.1 GI:22265139
DEFINITION QG19H22.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION QG19H22, mRNA sequence.
VERSION BO875918
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.
REFERENCE 1 (bases 1 to 622)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compositae.ucdavis.edu/
 Unpublished (2002)
JOURNAL Contact: Alexander Kozik [R.W.Michelmore]
COMMENT Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA.Contig6253, see http://cgdb.ucdavis.edu/
 for details.
Plate: QG19 row: H column: 22.
FEATURES Location/Qualifiers
 source
 1..622
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QG19H22"
 /clone_lib="QG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="vector: pBRCDNASFIAB; The library was constructed

from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transfections made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG_L1B-QC_ABCDI lettuce salinas
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=CTTGACGGG"

BASE COUNT 209 a 116 c 130 g 167 t
ORIGIN

Alignment Scores:

Pred. No.: 0.439 Length: 622
Score: 98.00 Matches: 47
Percent Similarity: 40.47% Conservative: 40
Best Local Similarity: 21.86% Mismatches: 92
Query Match: 3.09% Indels: 36
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ875918 (1-622)

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAAACGAGCAACAGCAAGGTACATCATCATCAAAAG-----GGTACTGAG 53
QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86
QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
Db 87 -----CTTGCTGGTAGTATTCATCACTGCTGATACACCTCAATCATCATCAACAAA 140
QY 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAAACTGAC-----ATGGCAAAATCAATGATTGATACACAGCACTCTTCTGCAAGGAG 194
QY 106 LysGlyTyrPheSerLysValSerLeuThrCluLysAspGlyAlaTyrThrValHis 125
Db 195 AATGGTATATTCGGTGTACCTTCACTTCAATTAACATCAGAAATGTTGGTGGATCAATCATCAA 254
QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATTCAGTACCGCAAGCATTAGAACACCAAGTCGGTTCGGAGGTATC 314
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
Db 315 CCAATTGTTGGGAATGCAGCAGAA-----TCTATTATTACACACAAACAGAC 362
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATGCGGTTGAAGCTCAACTTCGTCGGTCAAGAGCATGTTTATTAGGA 422
QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 423 GCTATTTTATCAGCGGCAGCAGGTTGTAGACAGATCTTGGAAACACA----- 470
QY 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAla 223
Db 471 -----GCAAAACAAATCATTTGAAAGGTTCTGATTCGTCTGTC 506
QY 224 PheGlyAspPheGluIleThrGlyThrGln-----ArgTyrPro 236
Db 507 GAAAAAACCTAGAGAAACACACATTCAGCATTTGGAGATATCCC 551

RESULT 24

BF535617

LOCUS

DEFINITION 602054012F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:419306 5', mRNA sequence.

ACCESSION

BF535617.1 GI:11622985

VERSION

EST.

KEYWORDS

house mouse.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 994)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-femail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM9525 row: g column: 03

High quality sequence stop: 656.

Location/Qualifiers

1. 994

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:419306"

/lab_host="NCI_CGAP_SG2"

/note="Organ: Salivary gland; Vector: pCMV-SPORT6; Site: 1"

dt. Average insert size 1.3 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 296 a 226 c 296 g 176 t

ORIGIN

Alignment Scores:

Pred. No.: 0.902 Length: 994
Score: 98.00 Matches: 76
Percent Similarity: 34.47% Conservative: 35
Best Local Similarity: 23.60% Mismatches: 124
Query Match: 3.09% Indels: 88
DB: 12 Gaps: 14

US-09-857-669-2 (1-609) x BF535617 (1-994)

QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
Db 5 AGGGTCGAGGAGCGGAGCGGAGCTTAGTCTCGGGAGGAGTGGAGCGCTGAGCGG 64
QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 65 TTAACGTCTACAGGCTTCCTGCTCCGCGCAGCTCTCGCGCGCGCGCTCTCTGAGACAC 124
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGln---Gln 164
Db 125 TCGGCTTCTCCCTGCTTGGCCAAAGAGGTAAACAACATGACTGACTGCTGCTCATATCT 184
QY 165 ProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 185 GCTCCTGGGGAGAAACCTGTGCAGCAACATGGGAGAAA-----CTACAT 229
QY 185 AlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsn 204
Db 230 CGACGCGCACCAAG-----AACATAATCTTTCGCTCTCT 265
QY 205 -----ProAspThrAlaThrValAspLeuAsnValValValAspSerGly 219
Db 266 TCCAACTTCAACATCTCTGACCTAAAGGTTCGCAGCTTGGATGCTTGGTT----- 316
QY 220 ArgProIleAlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIle 239


```

Db 1546 TCAGCACTGGGACCCAGC 1563
::: ||||| |||
RESULT 26
B0871709 630 bp mRNA linear EST 15-AUG-2002
LOCUS
DEFINITION
B0871709.1 yg.abl QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
QGI12J17, mRNA sequence.
ACCESSION
B0871709
VERSION
B0871709.1 GI:22258259
KEYWORDS
EST.
SOURCE
Lactuca sativa.
ORGANISM
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE
1 (bases 1 to 630)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
for details
Plate: QGI12 row: J column: 17.
FEATURES
source
Location/Qualifiers
1..630
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGI12J17"
/lab_host="E.coli"
/label="QG-ABCDI lettuce salinas"
/note="vector: pBRCDNASFIAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB-QG-ABCDI lettuce salinas
TAG_TISSUE-flowers pre-fertilized
TAG_SEQ-GCTTACGGG*
BASE COUNT 213 a 118 c 130 g 169 t
ORIGIN
Alignment Scores:
Pred. No.: 0.593 Length: 630
Score: 97.00 Matches: 45
Percent Similarity: 39.73% Conservative: 42
Best Local Similarity: 20.55% Mismatches: 88
Query Match: 3.06% Indels: 44
DB: 14 Gaps: 9
US-09-857-669-2 (1-609) x B0871709 (1-630)
Qy 26 LeuSerGluAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAACAGGACAGCAACAGGATCATCCATCCCAAAAG-----GGTACTGAG 53
Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65

```

```

Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86
|||
Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLys 85
|||
Db 87 -----GTTGTGGTAGTATTCCTCAATCATTTGGTGATCAGCTCAATCATCCACAAA 140
|||
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
|||
Db 141 CAAACTGAC-----ATGGCAATCAATGATTGATCCACGACTCTCTCTCAAGGAG 194
|||
Qy 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
|||
Db 195 AATGCTATATTCGGTGGTATCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 254
|||
Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
|||
Db 255 ATGGGTCTGATTCAGTACGACCAAGCATAGAAACACCAAGTCGGTTCGGAGGTATC 314
|||
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
|||
Db 315 CCAATTGTTGGGAATGCAGCAGAA-----TCTATTATTACAACACAAACAGAC 362
|||
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
|||
Db 363 ATGGCTAGTATGCGGTGTAAGAGCTCAACTTCGTCGGTCAAGAAGCATGGTTATTAGGA 422
|||
Qy 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
|||
Db 423 GCTATTTATCAGCGCAGGAGGAGTGTAGAGATGTTGGAACACACA----- 470
|||
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219
|||
Db 471 -----GCAACCAAAATCATTTGAAAGGTTCTGATCTCTCTCTCTCTCTCTCTCT 506
|||
Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
|||
Db 507 GAAAAACCCCTAGAGAAACCAATCCAGCATTCGAGATATCCCAATTCCTAGATCT 563
|||
RESULT 27
B0479447 690 bp mRNA linear EST 03-JUN-2002
LOCUS
DEFINITION
B0479447.1 y1 Strongyloides ratti PA female naive S11 TOPO v1
Strongyloides ratti cDNA similar to SW:YMT1.CAEP.L P98080
HYPOTHETICAL 51.7 KD PROTEIN F56D2.1 IN CHROMOSOME III. [1] ;, mRNA
sequence.
ACCESSION
B0479447
VERSION
B0479447.1 GI:21315414
KEYWORDS
EST.
SOURCE
Strongyloides ratti.
ORGANISM
Strongyloides ratti.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
REFERENCE
1 (bases 1 to 690)
McCarte,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarte JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Claire Murphy and Dr. James McCarter
at Washington University, St. Louis. Parasitic adult females were
collected from naive animals and provided by Dr. Mark Viney of

```

University of Bristol, Bristol, UK (Mark.Viney@bristol.ac.uk).
Seq primer: SL1 primer
High quality sequence stop: 525.

FEATURES

source
1..690
/organism="Strongyloides ratti"
/db_xref="taxon:34506"
/clone_lib="Strongyloides ratti PA female naive SL1 TOPO v1"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pCRII-TOPO (Invitrogen); Site_1: EcoRI; Site_2: EcoRI; The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. Oligo(dT)-SL1 PCR based library. cDNA PCR products of size >400 nucleotides containing SL1 on the 5' end and oligo(dT) on the 3' end were non-directionally cloned into pCRII-TOPO (Invitrogen) following the TOPO TA cloning protocol. Parasitic adult females were collected from naive animals and provided by Dr. Mark Viney of University of Bristol, Bristol, UK (Mark.Viney@bristol.ac.uk)."
BASE COUNT 239 a 102 c 123 g 226 t
ORIGIN

Alignment Scores:
Pred. No.: 0.785 Length: 690
Score: 96.50 Matches: 58
Percent Similarity: 38.89% Conservativeness: 33
Best Local Similarity: 24.79% Mismatches: 76
Query Match: 3.04% Indels: 67
DB: 14 Gaps: 12

US-09-857-669-2 (1-605) x BQ479447 (1-690)

Qy 18 HisAlaTyrAlaPro---AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeu 36
|||||:|||||: ||:|||||:|
Db 46 CATGCATTGCTCCAAATTCAGCTGATGTT----- 75

Qy 37 PheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPheProVal--- 55
|||:|||||:| ||:|||||:|
Db 76 ---TTGTCATCATCTAATCTTACTGAATCTACACATTTAAATAATGGAATTAAGTCGCT 132

Qy 56 -----Arg 56
|||
Db 133 GTTAATAATAATGGAGCTCAATGACAACCTGTTGCTGTTGGATTGATTCAGGATCAGCT 192

Qy 57 IleAspThrGlnAspSerGlu---IleLysAspMetValGluGluHisLeuProLeuIle 75
|||||:|||||:| ||:|||||:|
Db 193 TATGAACCTGAAGAAATAATGGAGTTCGTAATTTAGTTGAACATCTTCTTCACAAAGGA 252

Qy 76 ThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAla 95
|||:|||||:| ||:|||||:|
Db 233 ACTACAAAGATCAACAGTT----- 273

Qy 96 ProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu 115
|||:|||||:| ||:|||||:|
Db 274 -----CAATTAGACACAGAAATAGGTAAGATTGGTGTAGATTATCATCTCTAAGTACT 327

Qy 116 ThrGluLysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAla 135
|||:|||||:| ||:|||||:|
Db 328 AGAGAA-----CATACTCCCTATTTTGTCTCAAGCTCCATCCAGAAAGTTGAA 375

Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeu-----SerAspGlyAsn 150
|||:|||||:| ||:|||||:|
Db 376 CAAGTT---GTTGAATATATCTAGCTGATGTTTACGAAACTCTAAATTTAGTGAATCTAAT 432

Qy 151 LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPhe 170
|||:|||||:| ||:|||||:|
Db 433 ATT---GAAAGAGACAAATAATTTGTTAAATAAGTTGGATGAAGTTGAGAAATGATTAT 489

Qy 171 AspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGly 190
|||:|||||:| ||:|||||:|

Db 490 CAAGAGGTAGTTTGTATAAC-----CTTCATGCTACAGCTTTTCCAAGGA 534

Qy 191 TyrProLeuAlaLys-----LeuGlyAsnThrAlaAlaValAsnProAspThrAla 208
|||||:|||||:| ||:|||||:|

Db 535 ACACCATTTGGCTAAGAGTATTTATGGAAGCTACAGACAAATTCAAATCTTTTCACGTAAT 594
|||||:|||||:| ||:|||||:|

Qy 209 ThrValAspLeuAsnValValAspSerGlyArgProIle 222
|||||:|||||:| ||:|||||:|

Db 595 GATGTT---CTTAATTTTATGATCATCATTAACAACCTGTT 633
|||||:|||||:| ||:|||||:|

RESULT 28
BQ872562 577 bp mRNA linear EST 15-AUG-2002
LOCUS QG116A10.yg.abl QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION OGI16A10, mRNA sequence.
ACCESSION BQ872562
VERSION BQ872562.1 GI:22259121
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 577)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
P., Koltman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lal, Z.,
Church, S., Jackson, L. and Bradford, K.
TITLE Lactuca and Sunflower ESTs from the Compositae Genome Project
JOURNAL http://compgenomics.ucdavis.edu/
COMMENT Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmunsdon Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA-Contig6253, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG116 row: A column: 10.
FEATURES
source
1..577
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG116A10"
/clone_lib="QG-ABCDI lettuce salinas"
/lab_host="E.coli"
/note="Vector: pBRCNDNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB-QG-ABCDI lettuce salinas
TAG_TISSUE-flowers pre-fertilized
TAG_SEQ-GCTTGACGGG"

BASE COUNT 194 a 115 c 122 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 0.686 Length: 577
Score: 96.00 Matches: 45
Percent Similarity: 39.73% Conservativeness: 42
Best Local Similarity: 20.55% Mismatches: 88
Query Match: 3.03% Indels: 44
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0872562 (1-577)

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 Db 3 GTAAAGAGGACCAAGCAAGGATACATCCATCCAAAG-----GGTACTGAG 53
 Qy 46 SerValLysLeuLysProLysPheProValArgLeuAspThrGlnAspSerGluLeuLys 65
 Db 54 TCAGCCAGC-----CAAGCTTTAGACACAACAATACT----- 86
 Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
 Db 87 -----GTTGGTGGTAGTATCCCAATCATGTGTATACAGCTCATCATCATCCACAA 140
 Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
 Db 141 CAAACTGAC-----ATGGCAATCAATGATGATACACGACTTCTCTGTCAGGAG 194
 Qy 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AATGGTATATTCGGTGCATCATCTCATCTGATCAGATGTTGGTGGATCAATCATCCAA 254
 Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGTCTTCGATTCAGTCCGAGCATTTAGAAACCAACCAAGCTCGTTCGGAGGTATC 314
 Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 315 CCAATTGTTGGATGTCAGCAGAA-----TCTATTATTACACACACACACAGAC 362
 Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn----SerLysThrSerValLeuGly 184
 Db 363 ATGGCTAGTATCGGTTGAAAGCTCAACTTCGTGGTCAAGACATGGTTATTATAGA 422
 Qy 185 AlaVal----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
 Db 423 GCTATTTTATCAGCGCAGGAGGATGTTAGAGATGTTGGAAACACAC----- 470
 Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219
 Db 471 -----GCAACAAATCATTTGAAAGGTTCTGATTCTGTCTC 506
 Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
 Db 507 GAAAAAACCCTAGAGAAACACATCCAGCATTCGAGATATCCCAATTCATAGGATCT 563

RESULT 29
 BQ875018
 LOCUS
 DEFINITION BQ875018.1 yg.ab1 OG_ABCDI 660 bp mRNA linear EST 15-AUG-2002
 QG16020, mRNA sequence.
 ACCESSION BQ875018
 VERSION BQ875018.1 GI:22261578
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.
 1 (bases 1 to 660)
 REFERENCE
 AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., Van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, X., Lai, Z.,
 Church, S., Jackson, B. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 JOURNAL Unpublished (2002)
 COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmil.ucdavis.edu]
 belongs to contig QG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
 for details.
 Plate: QG16 row: 0 column: 20.

FEATURES
 source
 Location/Qualifiers
 1..660
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QG16020"
 /clone_lib="QG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_LIB-QG_ABCDI lettuce salinas
 TAG_TISSUE=flowers pre-fertilized
 TAG_SEQ=GGTTGACGGG"

BASE COUNT 222 a 121 c 136 g 180 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.844 Length: 660
 Score: 96.00 Matches: 46
 Percent Similarity: 40.57% Conservative: 40
 Best Local Similarity: 21.70% Mismatches: 96
 Query Match: 3.03% Indels: 30
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ875018 (1-660)

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 Db 3 GTAAAGAGGACCAAGCAAGGATACATCCATCCAAAG-----GGTACTGAG 53
 Qy 46 SerValLysLeuLysProLysPheProValArgLeuAspThrGlnAspSerGluLeuLys 65
 Db 54 TCAGCCAGC-----CAAGCTTTAGACACAACAATACT----- 86
 Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys-85
 Db 87 -----GTTGGTGGTAGTATCCCAATCATGTGTATACAGCTCAATCCATCATCCAAA 140
 Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
 Db 141 CAAACTGAC-----ATGGCAATCAATGATGATACACGACTTCTCTGTCAGGAG 194
 Qy 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AATGGTATATTCGGTGCATCATCTCATCTGATCAGATGTTGGTGGATCAATCATCCAA 254
 Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGGTTCTGATTCAGTCCGAGCATTTAGAAACCAACCAAGCTCGTTCGGAGGTATC 314
 Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 315 CCAATTGTTGGAAATGTCAGCAGAA-----TCTATTATTACACACACACAGAC 362
 Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn----SerLysThrSerValLeuGly 184
 Db 363 ATGGCTAGTATCGGTTGAAAGCTCAACTTCGTGGTCAAGACATGGTTATTATAGA 422
 Qy 185 AlaVal----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
 Db 423 GCTATTTTATCAGCGCAGGAGGTTGAGAGATGTTGGAAACACACACAAAAATC 482

```

QY 204 -----AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 483 ATTGAAAAGGTTCTGATCTGCGAAAAACCTAGAGANAACCATCCAGC----- 536

QY 221 ProileAlaPheGlyAspPheGluIleThrGlyThr 232
Db 537 -----ATTGGAGATATCCCAATTTAGGATCT 563

RESULT 30
BQ875030
LOCUS BQ875030 677 bp mRNA linear EST 15-AUG-2002
DEFINITION QG16P11.yg.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
QG16P11, mRNA sequence.
ACCESSION BQ875030
VERSION BQ875030.1 GI:22261590
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 677)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmall.ucdavis.edu]
belongs to contig QG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG16 row: P column: 11.
FEATURES
source
Location/Qualifiers
1..677
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG16P11"
/clone_lib="QG-ABCDI lettuce salinas"
/lab_host="E.coli"
/note="Vector: pBRcDNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB=QG-ABCDI lettuce salinas
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=GCCTGACGGG"
BASE COUNT 231 a 122 c 140 g 184 t
ORIGIN

Alignment Scores:
Pred. No.: 0.878 Length: 677
Score: 96.00 Matches: 45
Percent Similarity: 39.73% Conservative: 42
Best Local Similarity: 20.55% Mismatches: 88
Query Match: 3.03% Indels: 44
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ875030 (1-677)

```

```

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAAACGAGGCAACAGCAAGGTACATCCATCCATAAAG-----GGTACTGAG 53

QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86

QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
Db 87 -----GTTGTTGGTAGTATTCCAATCATGGTGATACAGCTCAATCCATCATCACCAA 140

QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAAACTGAC-----ATGCCAATCAATTTGGTTGATACCACGACTTCTTCTGTCAGGAG 194

QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaIleThrValHis 125
Db 195 AATGGTATATTCGGTGTGCTATCATTCTGACTCAGAAATCTTGGTGCATCAATCATCCAG 254

QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATTCTAGCTAGCCCAAGCATTTAGAAACAACCAAGTCGGTTGTGGAGGTATC 314

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 315 CCAATTTGTTGGAAATCCAGCAGAA-----TCTATTATTACAACACACACAGAC 362

QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATTCGGTGTGAAAGCTCAACTTCGTCGGTCAAGAAGCATGTTTATTAGGA 422

QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 423 GCTATTATTATCAGCACCGACAGGTTGTAGAGAGATGTTGGAAACACA----- 470

QY 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
Db 471 -----GCACACAAATCATTTGAAAAGGTTCTGATTCTGTC 506

QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 507 GAAAAACCCCTAGAGAAAACCATCCAGCATTTGGAGATATCCCAATTTAGATCT 563

RESULT 31
BQ874536
LOCUS BQ874536 731 bp mRNA linear EST 15-AUG-2002
DEFINITION QG15M18.yg.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
QG15M18, mRNA sequence.
ACCESSION BQ874536
VERSION BQ874536.1 GI:22261177
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
REFERENCE 1 (bases 1 to 731)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

```


DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0850925 (1-735)

Qy 26 LeuSerGluAsnLysAlaLaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 Db 3 GTAAACGAGGCACACAGAGTACATCCATCCAAAG-----GGTACTGAG 53

Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLys 65
 Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86

Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLys 85
 Db 87 -----GTTCTGCTAGTATCCCAATCTGATACAGCTCAATCCCATCCACAAA 140

Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
 Db 141 CAAACTGAC-----ATGGCAATCAATGATGATACACGACTCTTCTGTCAAGGAG 194

Qy 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AATGGTATATTCGGTCTCATCTTCACTCAGAACTGTTGGTGCATCAATCCAAA 254

Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGGTTCTGATTCAGCTAGCAAGCATTAGAAACAAACCAAGTCGGTTCGGAGGTATC 314

Qy 146 LeuSerAspLysAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
 Db 315 CCAATTTGTTGGGAATCGACAGAA-----TCTATTATTACAACACAAACAGAC 362

Qy 166 ValGlySerPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
 Db 363 ATGGCTAGTATCGGTTGAAAGCTCAACTTCGTCGTCAGAACGATGGTTTATTAGGA 422

Qy 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
 Db 423 GCTATTTTATCAGCGGCAGCAGGTTGTAGAAGATGTTGGAACACACAGCAATCAAATC 482

Qy 204 -----AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
 Db 483 ATTGAAAGAGTTCTGATCTGCGAAAAAACCCCTAGAGAAACACCATCCAGC----- 536

Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
 Db 537 -----ATTGGAGATATCCCAATCTTAGGATCT 563

RESULT 33
 LOCUS B0871570
 DEFINITION B0871570.1 yg.abi OG_ABCDI lettuce salinas Lactuca sativa cdna clone
 ACCESSION B0871570
 VERSION B0871570.1 GI:22258120
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE 1 (bases 1 to 757)
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevaller,P., Ziegler,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
 TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
 JOURNAL http://comgenomics.ucdavis.edu/
 COMMENT Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)

```

Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
Db 501 -----GCAACAAATCATTTGAAAGGTTCTGATTCGTCTGTC 536

Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 537 GAAAAAACCTTAGAAAAACACATCCAGCATTCGAGATATCCCAATTCTAGGATCT 593

RESULT 34
BM919327 1072 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT_6715791 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5748568
5', mRNA sequence.
ACCESSION BM919327
VERSION BM919327.1 GI:19369706
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1072)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12776 row: 1 column: 17
High quality sequence stop: 572.
Location/Qualifiers
1..1072
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:5748568
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dr primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT 270 a 340 c 292 g 170 t
ORIGIN

Alignment Scores:
Pred. No.: 1.78 Length: 1072
Score: 36.00 Matches: 79
Percent Similarity: 33.91% Conservative: 39
Best Local Similarity: 22.70% Mismatches: 116
Query Match: 3.03% Indels: 115
DB: 14 Gaps: 16

US-09-857-669-2 (1-609) x BM919327 (1-1072)

Qy 5 ProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyraProAla 24
Db 6 CCAAGCGCTTCTCCGACCAGGGAAG-----CCCCAC-----CCACGAGAGCC 50

Qy 25 AspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThr 44
Db 51 AAGATGTCCAGCAGCGGGCCAAAGCCACCAAGAGCGGCCACAGCGGCCACCA 110

```

```

Qy 45 GluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIle 64
Db 111 TCCATGTC-----TTCCAAATGTTTGACCACTCCAGATCCAGAGTTT 155

Qy 65 LysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAsp 84
Db 156 AAG-----GAGGCTTTCACACATGATTGACCAAGAACCGTGATGGCTTCATTGAC 203

Qy 85 LysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArg 104
Db 204 AAGGAG-----GACCTGCACGACATGCTGGCC 230

Qy 105 SerLysGly-----TyrPheSerSerLysValSerLeuThrGluLys 118
Db 231 TCGCTGGGGAAGAACCCACACAGCAATACCTGGAGGCGCATGATCAGC-----GAGGC- 283

Qy 119 AspGlyAlaTyThrValHisIleThrPro----- 128
Db 284 CCCGGGGCCCATCAACTTCACCATGTTCTCCATCACCATGTTTGGGGAGAGCTGAACGGCAC 343

Qy 129 GlyProArg-ThrLysIleAlaAsnValGly----- 138
Db 344 GGACCCCGAGGATGTGATTCGCAACGCTTTGCGCTGCTTCGACGAGGAAGCTCAGGTTT 403

Qy 139 -----ValAlaIleLeuGlyAspIleLeuSerAspG1 149
Db 404 CATCATGAGGACCACTCCCGGAGCTCTCACCACCATGGGTGACCGCTTCACAGATGA 463

Qy 149 yAsnLeuAlaGluTyTyraArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAs 169
Db 464 GGAAGTGGACGAGATGTACCGGGAGGCA-----CCCATTTGATAAGAA 505

Qy 169 pPheAsp-GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgL 189
Db 506 AGGCAACTTCAACTACGTACGTGGAGTTACCCCGCATCTCAACATGCGC----- 552

Qy 189 yGlyTyProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla- 208
Db 553 -----CCAAGGATAAGAGACTAGGGCACCCCGCCCGCCCTGCACCCCGCC 598

Qy 209 -----ThrValAspLeuAsnValVal 216
Db 599 CCCCGCGCAGTCACCCCTCCCGGCACACACCGCTCCATACCACTCCCTGCGCCATGAGCC 658

Qy 216 aLaspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyP 236
Db 659 TCGCTCAGGGATCCCGCC-----TTTGAGGGGTAGGGTCCCGAGTTCCCGAG 703

Qy 236 roGluGlnIleValSerGlyLeuAlaArgPhe----- 246
Db 704 TGGAAAAAACAGGGCCAGGGGNAAGTGGTGGCCCAAGCTTGAGGCAGATGTTTCCACAG 763

Qy 247 --GlnProGlyThrProTyraAspLeuLeuLeuAspPheGln-----G 262
Db 764 GTGACCCCGCAGAGCCCTGGGGCTATAGTTCTCTCACCCTTCCCAAGGAAAGACACCA 823

Qy 262 InAlaLeuGluGlnAsnGlyHisTyraSerGlyAlaSerValGlnAlaAspPheAspArgL 282
Db 824 CTTTTCCTTTGGGGACATCGGGCGCTGGGAAGGGGCGGAGAAC-----TTTAAAGGCA 874

Qy 282 euGlnGlyAspArgValPro 288
Db 875 CCCAAGGGGAAAGGGCCCCC 894

RESULT 35
BM975465 963 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8203637 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255749
DEFINITION 5', mRNA sequence.
ACCESSION BM975465
VERSION BM975465.1 GI:21786299
KEYWORDS EST.
SOURCE human.

```



```

/lab_host="DH10B"
/notes="vector: pDNR-LIB"
BASE COUNT 751 a 484 c 500 g 731 t
ORIGIN

Alignment Scores:
Pred. No.: 7.38 Length: 2466
Score: 95.50 Matches: 81
Percent Similarity: 36.44% Conservative: 56
Best Local Similarity: 21.54% Mismatches: 125
Query Match: 3.01% Indels: 114
DB: 11 Gaps: 20

us-09-857-669-2 (1-609) x BC013324 (1-2466)

Qy 74 LeuIleThrGlnGlnGluValLeuAspLys-----GluClnThrGly 89
Db 418 ATTGTGAATACAGCAATGAAGTCTCACAGATAAGATCTGTGTCCAGAAAGTATCTTC 477
Qy 90 PheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPhe 109
Db 478 ATGCTTGCA-----GGCCCTGATTGAAGGAGATCTCCAGAACTTTAAAGGGTCCCGA 531
Qy 110 SerSerLysValSerLeuThrGlu-----LysAspGlyAlaTyrThrValHisIle 126
Db 532 ACAGGTGAAGCAAAATTCACAAAGAGTCAATCTAGCTGCCCGGTTTCATCATTCAC--- 588
Qy 127 ThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeu 146
Db 589 ACAGTGGGACCTAAATATAAAGCGCTATCCGACAGCAGCT----- 630
Qy 147 SerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProVal 166
Db 631 ---GAGAGTTCCTTTATAGCTGTACAGAAACGTACTT----- 666
Qy 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaVal 186
Db 667 -----CAACTAGCAAAAGAGCAGTCAATGTCTCTGTGGTCTCTGTCTGTC 711
Qy 187 -----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 712 ATCAATTCGCAAAAGCGGGTATCTCTTA----- 741
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValVal-ValAspSer-----G1 219
Db 742 ---GAGAGTGCACACACATACCACTTCGCACCTGTAAAGAGATTCCTAGAGATTCATGG 798
Qy 219 YargProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnI 239
Db 799 GAAACCATTT-----GAAAAAGT 816
Qy 239 eValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAs 259
Db 817 AGTATTTGCTGCTCTGATCTTGAAGAGGGTACT---TACCAAAAGCTGCCTACCTCTCTA 873
Qy 259 pPheGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPh 279
Db 874 CTTCACAGGTCAATAAAGAGGAGAAATCGA---TCATTGCCCTACCTACCTGCAGATAT 930
Qy 279 eAspArgLeuGlnGlyAsp-----ArgValProValLysVa 291
Db 931 TGGAAATGCAGAGGGGCGCTGTGGTACCTGAAGCAGACAGATAGATAAGTGAGAAACC 990
Qy 291 LserValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTy 311
Db 991 TGGTCTCCAGAGATACCAAGAGAGAGATGAAGCGTTGGAGTTGGAT----- 1042
Qy 311 rGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleG1 331
Db 1043 -----CTCTCTTTTCATTGG 1056
Qy 331 ySerValValTrp-----AspMetAspLysTyrGluThrThrLeuAlaAl 346

```

```

Db 1057 CTCCTCATGCTTTTGTCTCGAATGGAAGGAGATATTGACAGCAAGAAACAACTGATCTTCA 1116
Qy 346 agly---IleSerGln-----ProArgAsnTyrArgGlyAsnTy 358
Db 1117 GGGCAATATATCAGAGCGAGCTCTGCAGAGCAGCATCAAGAAATAT-----AATCG 1170
Qy 358 rTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPh 378
Db 1171 CTGGTTATGT-----CAAGCAAGATCTCAGGATCTGTCTGATATTTGCTTC 1215
Qy 378 eSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaG1 398
Db 1216 TCTAAAGCCTTATAC-----CAACACAGGTCTTGATACTGTGGTGCAGACAGT 1263
Qy 398 upheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 1264 GATGCTGTAGTTGGAGAAACATTCCTGTAAACATTAATAGATATG 1309

RESULT 38
B0875852
LOCUS B0875852
DEFINITION B0875852 578 bp mRNA linear EST 15-AUG-2002
ACCESSION QG19E21.yg.ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
VERSION QG19E21.yg.ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
KEYWORDS B0875852.1 GI:22265073
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa.
REFERENCE 1 (bases 1 to 578)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
P., Kollman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmudson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig6253, see http://cgdb.ucdavis.edu/
for details.
Plate: QG19 row: E column: 21.
FEATURES
Location/Qualifiers
source 1..578
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG19E21"
/clone_lib="OG_ABCDI lettuce salinas"
/lab_host="E.Coli"
/notes="Vector: pBRCNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG_LIB=OG_ABCDI lettuce salinas
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=CGTTGACGG"
BASE COUNT 196 a 115 c 121 g 146 t
ORIGIN

```


Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
 ... ||||| ... ||| ... ||| ... |||||
 Db 363 ATGGCTAGTATCGCGTTGAAAGCTCAACTTCGTCGGTCAAGAACGATGGTTTATTAGGA 422

Qy 185 AlaVal-----ThrArgGlyGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
 ||||| ... ||| ... ||| ... |||||
 Db 423 GCTATTTTATCAGCGCGCAGGAGCTTGTAGAAGATGTTGGAACACA----- 470

Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219
 ... ||||| ... ||| ... ||| ... |||||
 Db 471 -----GCAACAAAATCATTTGAAAAGGTTCTGATTTCTGTC 506

Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
 ... ||||| ... ||| ... ||| ... |||||
 Db 507 GAAAAACCTAGAGAAAACCATCCAGCATTTGGAGATATCCCAATTTCTAGGATCT 563

RESULT 40
 B0874935
 LOCUS
 DEFINITION B0874935 622 bp mRNA linear EST 15-AUG-2002
 OGI6K24.Y9.ab1 OG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 B0874935
 ACCESSION B0874935
 VERSION B0874935.1 GI:22261495
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Lactuca sativa.
 Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE
 AUTHORS
 1 (bases 1 to 622)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J., Ellison
 , P., Kolkman, J., Stabaugh, M.S., Livingston, K., Zhou, Y., Lal, Z.,
 Church, S., Jackson, B. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmunsdon Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
 for details.

FEATURES
 source
 Plate: QG16 row: K column: 24.
 Location/Qualifiers
 1..622
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QGI6K24"
 /clone_lib="OG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_L1B-OG_ABCDI lettuce salinas
 TAG_TISSUE=flowers pre-fertilized
 TAG_SEQ=GCCTGACGGG"

BASE COUNT 210 a 116 c 129 g 167 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.02 Length: 622
 Score: 95.00 Matches: 45
 Percent Similarity: 39.73% Conservative: 42
 Best Local Similarity: 20.55% Mismatches: 88
 Query Match: 2.99% Indels: 44
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ874935 (1-622)

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 ... ||||| ... ||| ... ||| ... |||||
 Db 3 GTAAACGAGGCACACAGCAAGGTAGTATCCATCATCCAAAG-----GGTACTGAG 53

Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
 ||| ... ||||| ... |||
 Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAAATCT----- 86

Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
 ||| ... ||||| ... |||
 Db 87 -----GTTGTTGGTAGTATTCCAATCATTTGTATGATACAGCTCAATCCATCATCACCAAA 140

Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
 ... ||| ... ||| ... |||
 Db 141 CAACACTGAC-----ATGGCAATCAATTGATTGATACACGACTTCTTCTGTCAGGAG 194

Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 ||| ... ||||| ... |||
 Db 195 AATGATATTCGGTGCTATCATCTCATTCATGATCAGATGTTGGTGATCAATCATCCNA 254

Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 ... ||| ... ||| ... |||
 Db 255 ATGGGTTCTGATTACAGTACAGCAAGCATTAGAACCAACCAAGCTGTTGTCGAGGTATC 314

Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 ... ||||| ... |||||
 Db 315 CCAATTGTTGGAAATGCACAGAA-----TCTATTATTACACACAAACACAGAC 362

Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
 ... ||||| ... ||| ... ||| ... |||||
 Db 363 ATGGCTAGTATCGCGTTGAAAGCTCAACTTCGTCGGTCAAGACGATGTTTATTAGGA 422

Qy 185 AlaVal-----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
 ... ||| ... ||| ... |||
 Db 423 GCTATTTTATCAGCGCGCAGGAGCTTGTAGAAGATGTTGGAACACA----- 470

Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219
 ... ||| ... ||| ... |||
 Db 471 -----GCAACAAAATCATTTGAAAAGGTTCTGATTTCTGTC 506

Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
 ... ||||| ... ||| ... ||| ... |||||
 Db 507 GAAAAACCTAGAGAAAACCATCCAGCATTTGGAGATATCCCAATTTCTAGGATCT 563

Search completed: November 9, 2002, 02:55:17
 Job time : 1979 secs

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	146	4.6	3472	9	US-10-027-806-4		Sequence 4, Appli
2	122	3.8	1781	9	US-09-995-749A-2		Sequence 2, Appli
3	118.5	3.7	815	10	US-09-815-242-5106		Sequence 5106, Ap
4	116	3.7	922	10	US-09-886-468-19		Sequence 19, Appl
5	111	3.5	976	10	US-09-871-212-6		Sequence 6, Appli
6	108.5	3.4	794	10	US-09-864-761-37663		Sequence 37663, A
7	108.5	3.4	1375	10	US-09-740-274-4		Sequence 4, Appli
8	108.5	3.4	3158	10	US-09-815-242-12611		Sequence 12611, A
9	106	3.3	3034	10	US-09-737-149-25		Sequence 25, Appl
10	106	3.3	3034	10	US-09-737-149-30		Sequence 30, Appl
11	104.5	3.3	382	10	US-09-912-020-299		Sequence 299, App
12	102	3.2	2025	10	US-09-815-242-5703		Sequence 5703, Ap
13	101	3.2	982	10	US-09-841-132-176		Sequence 176, App
14	101	3.2	1006	10	US-09-841-132-176		Sequence 190, App
15	100.5	3.2	939	10	US-09-815-242-10797		Sequence 10797, A
16	100	3.2	2086	10	US-09-815-242-5639		Sequence 5639, Ap
17	100	3.2	5795	10	US-09-815-242-12610		Sequence 12610, A
18	99.5	3.1	1025	10	US-09-897-056-7		Sequence 7, Appli
19	99	3.1	764	10	US-09-747-521-4		Sequence 4, Appli

Qy 335 WDMKYETTTLAAGISQPNRYGNWTS----- 361
Db 2290 -DLEP-EDVIPGISFSDGTSLFMTGEDVDHIHEYALNEPMDIRNAILAGLSLSAVNG 2347
Qy 362 -----NVSYNSTTONLEKRAFSGGIWYVRDRAGI-----DARLGAEFLAEG 403
Db 2348 APRGLDISEDGTTHAHTMRGRDFGT-----PASLVNHLPGQYSLTUDAPAFAP-YPVEE 2400
Qy 404 RKIPGSDIDLGNHSHATMLTASWKKQLLN-NVLHP-ENG-HYLDGKI-----GTTILGTF 453
Db 2401 EGARG-DLAFSDGDMRFVAGVNNHRLQYNLLSPDYTENAEHFISTDLTLADRGTGLVFP 2459
Qy 454 LSSTALIRTSARAGY-----FFTPENKLGTFIIRQOAGYTVARDNADVPSPGLMFRSGGASS 510
Db 2460 SDENDFESTGARAQFVRQFTNRPDASTITLSDNGLYKVSVDG-LPSGIRFTPDGMKM 2517
Qy 511 -VRGYELDSICLAGPNGSVLPERALLVGSLEYQLPFTRTLSGAVFHDGMDAAANFKRML 569
Db 2518 FISGOETAMI-----YQYSLPSYDTSGAVRDRVEIVAGLFRNAGL 2558
Qy 570 K-----HCSGLGVWRFSPPLAPFSF-DIAYGHSDDKKIRWHISLG 606
Db 2559 SVGLNEPSPGDFSDGMELYVTGSLGVHRYFLP-SPYGLEDAAYGGSFHTPRESTPLG 2617

RESULT 2
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 3.8%; Score 122; DB 9; Length 1781;
Best Local Similarity 18.1%; Pred. No. 0.088;
Matches 128; Conservative 85; Mismatches 231; Indels 262; Gaps 34;

Qy 35 ALFNKSPDTSVKLPKFPVRIDTQSEIKDMVBEHLPLTQOOEEVLDEKQTFGLAE 94
Db 1109 SLADRENNSTENVVI-PNYSF-VRAHDNNSODQIONAI-----RDVTGDKYHTFTFED 1159
Qy 95 APDNNVTKMLRSKGVFSSKVSLETGDAYTVHITPG-----PRTKIANVGVAIIGDILSDG-- 149
Db 1160 EOKGIDAYIQDNSTVTKRYNDINIPASYAILLTNKTIPR-----VYVGDLTYDGGQ 1211
Qy 150 ---NLAEY-----RNALENMQQPVGSDFDODS-WEN 177
Db 1212 YMEHTRYDYFLTNLLKSRVKYVAGGOSQPMVSGNNNLTLSYRKGAMATATDTGTDE 1271
Qy 178 SKTSVLGAVTRKGYPLAKLG-----NTRAAVNPDATATVDLVNVDVSGRP 221
Db 1272 TRTOGIGVVV-SNTPNLKLGVDNKKVLMHGAANKQYRAAVALTTTDCV-INYTSQCAP 1329
Qy 222 IAF---GDFEI-----TGTORYPEQIVSGL-----ARQPG 249
Db 1330 VAMTDENGDLTSLSHNLVNVNGKEADTAVQGYANPDVSGYLANVWVPGASDNQDARTAPS 1389

Qy 250 TPYDLILLDFQALEQNGHYSGASVOADFDR---LQGDRVPVKVSVTEVRHKLKLETGIR 306
Db 1390 T-----EKNSGNSAYRTNAAFDSNVIFFAFSNFVYTPTKESERANVRIAQN 1435
Qy 307 LDSYGLG---GKIAYDYNLFNKGYIGSVVMD---MDKYETTTLAAGISQPNRYGNV- 358
Db 1436 ADFFASLGFTSFEMAPOYNSKDRTEFLDSTIDNGYAFTRDYD---LCMSEPNKYGTDED 1491
Qy 359 -----WTSNVSYNR-----SITONLEKRAFSGGIW-----YVR 386
Db 1492 LRNAIQALHAKAGLVMDWPDQIYNLPCKEVATVTRVDDR---GNVMKDAIINNLYV 1548
Qy 387 DRAG---IDARLGAELAEGRKI-----PGSIDLGNHSHATMLTASWKKQLLN- 431
Db 1549 NTICGGYQKKYGGAFDLKLOKLYPEITFTKKQVSTGVAID-----PSOKITEWSAKYPNG 1603
Qy 432 -NVLHPENGHYL--DG-----KIGTTLGTFLS----- 455
Db 1604 TNLHRSQYVVLKADGGQYVNLGTTTKQFLPIQLTGEKKQCNQGVKNGDGNYYFYDLAG 1663
Qy 456 ---STALIRTSARAGYFTFPENKKL-----GTFIIRGOAGYTVARDNADVP 498
Db 1664 NMVNTTFIEDSVGNWYFFDQDKMVENKNHFYDVDSYGEKGTFF-----LK 1709
Qy 499 SGLMFRSGGASSVRG-YELDSIGLAGPNGSVLPERALLVGSLEYQL 543
Db 1710 NGVSFRGGLVOTDNGTYFFDNY-----GKMVRNQTINAGAMIYTL 1749

RESULT 3
US-09-815-242-5106
; Sequence 5106, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5106
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5106

Query Match 3.7%; Score 118.5; DB 10; Length 815;
Best Local Similarity 19.5%; Pred. No. 0.052;
Matches 112; Conservative 68; Mismatches 192; Indels 201; Gaps 30;

APPLICANT: Zhang, Linong
APPLICANT: Wu, Qiaohua
TITLE OF INVENTION: MODIFIED BOVINE ADENOVIRUS HAVING
TITLE OF INVENTION: ALTERED TROPISM
FILE REFERENCE: 293102003000
CURRENT APPLICATION NUMBER: US/09/871.212
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/208,678
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 976
TYPE: PRT
ORGANISM: Bovine adenovirus 3
US-09-871-212-6

Query Match 3.58; Score 111; DB 10; Length 976;
Best Local Similarity 22.24; Pred. No. 0.31;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;
QY 3 IKPTALL----LPALFFPPHAYAPAADLSENKAAGFALFNKNSPDTSVKLKPKP----- 53
DB 66 INPAGLLETNDLASAVFPPLASDAGNVTLNMSDGLYTKNKL-----AVKVGPGLSLDSN 121
QY 54 -PVRIDTQD-----SEIKDMVEEHLPLITQOOEEVLDKEQTGFLAEAPDNVYKTMRLRSKG 108
DB 122 NALQVHTGDGLTVTDKRVSLNTQAPLSTTSAGLSLLGLPSLHGEERLTVNT----- 174
QY 109 FSSKVSILTEKDCATVHTPGPRKIANVGVAIIGDILSDGNLAIEYRNALENQOQVGS 168
DB 175 ---GAGLQISNALAVKVGSGITVDAQNQLAASIGDGL----- 209
QY 169 DFDDDSWENSKTSVLG---AVTRKGYPLAKIGNTRAIVNPDATVDLNVVVDVSGRPIAFG 225
DB 210 ---ESRDKNVTWAGCLTITNQLTVA-TGN-GLQVNPE-GQLQLNITAGQGLNFANN 262
QY 226 DFEITGTQRYPEQIVSGILAREPQGT-----PYD-LDL-----LLDFOQAL 264
DB 263 SLAV-----ELGSSL-HFPFGQNVSLYPGDIIDRONRVTVPAGPGLRMLNHQAV 313
QY 265 EQNGHYSQASVQADFRLQGRVVPKVSVT---EVKHKLEGTGLRDS-----EYGLGG 315
DB 314 A-----SGDGLGVHSDTL---RLKLSHGLTFENGAVRAKLGPGGLGTDGSRVYVTRGR 365
QY 316 KIAYDYNNLNF-KG-YIGSVVWMDKYEITTLAAGISOPRNVYRGNYWTSNVSYNRSSTQNL 373
DB 366 RVANGQVIFSGRTAGT-----DSSLTINIRAPLQPSGPAITASLOGSGPITYNS 417
QY 374 EKRAF---SGGIWYVRDRAGIDARLGAEF-----LAEGRKIPGSDIDLG-----NS 416
DB 418 NNGTFGLSIGPGMWVDQNRQLQVNPAGLVFQGNLVPNLADPLAISDSKISLSLGPGLTQ 477
QY 417 HATMLTASWKQQLNNVLPENGYHLOGKIGTTLCTLSLTALTRTSARAGYFTTPENKK 476
DB 478 ASNALTLT-----LGNGLEFSN-QAVAKAGRGL-REFSSQALESSLITVGNGLT----- 525
QY 477 LGTFIRGQAGYTV-ARDN---ADVPGLMFRSGGASVVRGYELDSGLAGPNSVLP-- 530
DB 526 LTDVIRPNLGDGLVEVRDNKIIVKLGANLRPENGAVTA-----GTVNPSAPEAPPT 576
QY 531 ---BRALLVGSLEYQLPFTRTLTSAGVPHD-----MGDA-AANFKRMLKHGSGLVGR 578
DB 577 LTAEPPLRASNSHLQLSLS---EGLVVHNNALALQLGDGMVNHQGLTLRVGSGQLQR 631

RESULT 6
US-09-864-761-37663
Sequence 37663, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 37663
LENGTH: 794
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005366.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: P33450, EVALUE 5.00e-61
OTHER INFORMATION: EST_HUMAN HIT: BF529515.1, EVALUE 7.00e-78
US-09-864-761-37663

Query Match 3.43; Score 108.5; DB 10; Length 794;
Best Local Similarity 19.84; Pred. No. 0.36;
Matches 135; Conservative 75; Mismatches 184; Indels 287; Gaps 35;
QY 25 DLSNKAAG--FALFNKNSPDTSVKLK-----PKFPVRIDTQDSEIKDMVEEHLPLI 75
DB 81 EISEAVAPGTRFPLESADHPDVGNSLQTYELSRNEVFALRVQTRDSTK-----YAEIV 135
QY 76 TQOOEEVLDREQ-----TCFLAEAPDNVKTN-----LRSKGYFSSKVS 114


```
Db 136 L-----ERALDRERPSQLVLTALDGGTPALSASLP1HIKVLKDANDNAPVENQSLYRARVL 192
QY 115 LTEROGAYTHI-----TPGPRTKI-----ANGVAILGDILS-----D 148
Db 193 EDAPSGTRVVOVLATLDLDEGNGEIIYFSGSHNRAGVRQLFALDLVGTGMLTIKGRDLDFD 252
QY 149 GNLAEYRNALENMQQPVGS-----DFDODSWNSKTSVLGAVTRKGYPLAKLGN 199
Db 253 TKLHEIYIOAKDKGANPEGAHCKVLVEVDVNDNAPITVTSYSPVED-----APLGT 308
QY 200 RAAVNPTATVDLNVVDSGRPTAFGDFEITGTQRYEQVLSGLARFQ---PGTPYDLDLL 257
Db 309 IALL-----SVTDL-----DAGE-----NGLVTCVPPGLPFLSTSS 340
QY 258 LDFQALQEQNGHVSASVQADFRLQGRVP-----VKSYSYTEV 296
Db 341 LK-----NYFTLTSADLOR---ETVPEYNLSITARDAGTPPSLSALTIVRVQVSDI 388
QY 297 KRHKLETGIRLDSEYGLGGKIADYY-----NLFNKGYIGSVVMDMKRYETTLAAGISQPR 352
Db 389 NDNPPOS-----SOSYDVYIEENLPCAPILNLSVWDPD-----APQ 426
QY 353 NYRGNYSVNSVNRSTQMLERAFSG---GIWYVRDR-AGIDARLGAELAEGRKIPGS 409
Db 427 NARLSFEL-----LEOGAETGLVGRYFTINRDNIGVSSL-----VP-- 462
QY 410 DIDLGNHATMLTASWKROLLNNVLPENGHYLDGKGTGTLCTFLSSTALIRTSARAGYF 469
Db 463 -LDYEDRREFELTA-----HISDG-----CTPVLATNISVNI 494
QY 470 FTPENKLGTFIIRGQAGYTVARDNADVPGLMFRSGGA-----SSVRGY 514
Db 495 VTDNR-----DNA---PQVLYPRPGSSVEMLPRTSAGHLVSRVVGW 534
QY 515 ELD-----SIGLAG-PNGSVL-----PERALLY---GSLEYQL 543
Db 535 DADAGHNAMLSYLSLSPNSLSFAIGLHTGQISTARPVQDTPSROTTLVLKDNQGP 594
QY 544 PFTRTLSCAVPHDMGDAANF 564
Db 595 STTATLTVSVTEDSPEARAEF 615

RESULT 7
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
```

```
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 3.4%; Score 108.5; DB 10; Length 1375;
Best Local Similarity 19.7%; Pred. No. 0.86;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;

QY 30 KAAGFALFKNKSPTDESVKLKPKFPVPRIDTQDSEIKDMVEH-----LPLITQOQEEV-- 82
Db 729 RTSCVAVIEGNNS-----LRLKASDRVVVNMGAHKNQAYRPLLTITDNGIKA 777
QY 83 --LDKEOTGFL--AEEAPDNVKTMLRSKGYFSSKVS-----LTEKDAYTVHIT 127
Db 778 YHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASTA 837
QY 128 PGPRTKIANYGVAIIGDILSDG-----NLAERYRNA-----LENNQQPVGSDFDODS 174
Db 838 PSTDGSKVHQAALDSDSMFEGFSNFQAFATKKEEYTNVVIKNDKFAEANGVTFDEMAP 897
QY 175 WENSKT--SVLGAVTRKGYPLAKLGNTRAAVNPDATVDLNVVDSGRPIAFG----- 225
Db 898 QYVSSDTGSLDSVIONGYAP-----TDYDGLI-----SKPNKYGTADDLVK 940
QY 226 -----DFEITGTQRYE-QIVSGLARFQGTG-----YDLDLLDF 260
Db 941 AIKALHSKGIKVMADWVPDQMYALPEKEVVTATRVKDYGTVPAGSQIKNTLYVVD--GK 997
QY 261 QQALEQNGHVSQA---SVQADFRLQGRVPKVSYTEVVKRHKLETGIRLDSEYGLGKI 317
Db 998 SSGKDOAKYGGAFLELOAKYPEL-----FARKQISTGVPMDPSPVKIKOWS 1044
QY 318 A--YDYYNLFNKGYIGSVVWDM--DKYETTLAAGISOPRN-YRGNYSWTN----- 362
Db 1045 AKFPGNTNILGRG-AGVVLKDAQNTYFSLVSDNTFLPKSLVNPNGHTSSSTVGLVFDGK 1103
QY 363 --VSYNRSTTONLEKRAF--SGGIWYVRDR-----AGIDARLGAELAEGRKIPGSDI 411
Db 1104 GYVYVY--STSGNAKNAFISLGNWYVFDNNGYVMTGAQISNGANYFVLSNGIQLRNAIY 1161
QY 412 DLGNHATMLTASWKROLLNNVLPENGHYLDGKGTGTLCTFLSSTALIRTSARAGYFT 471
Db 1162 DNGKNKLSYYGNDGRY-----ENGYLFGQOOWRYFONGIMAVGLTRVHGAVQYFDA 1213
QY 472 PENKKLGTFIIR--GOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLA----- 522
Db 1214 SGFOAKGQFITTAGDKLRY-FDRSGNQISNRFVRN---SKGEWFLFDHNGVATGVTTF 1269
QY 523 -----GPNGSVLPERALLY---GSLEYQLP-----FTRTLGSAVF-HDMGDAA 561
Db 1270 NGQRLYFKPNG--VQAKGEFIRDANGYLRYDDPNSGNEVRNRFVRNSKGEWFLFDHNGIA 1327
QY 562 ANFRMKLKHGSLGVVRWF 580
Db 1328 VTGARVYVNGHASILSMVF 1346

RESULT 8
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match 3.48; Score 108.5; DB 10; Length 3158;
Best Local Similarity 20.2%; Pred. No. 3.2;
Matches 91; Conservative 66; Mismatches 159; Indels 135; Gaps 20;

Qy 27 SENKAAGFALFKPKSPDESVKLPKFPVRIDT-QDSEIKDMVEHLPLITQQQ----- 79
Db 1004 SLNTAMGNLI--NAIADHQAEVQGRNFP-INADTKQATYTTAVNAEAMINKQTCQANQ 1060

Qy 80 ---BEVLKDEQGTGLAEAPDNVK-----TMLRSKGYFSSKVSLSLTKDCGAYT 123
Db 1061 PEVEQATIKVQTTLQALNGDHNQVAKTNATQAI DALTSLNDPQK TALKDQVTAATLVTA 1120

Qy 124 VH-ITPGPRTKIANGVAILGDIISDGNLA EYRNALENWQPVGSPFDQDSWENSKTSV 182
Db 1121 VHQTEQNANT--LNOAMHGLRESIQD-NAATKANSKYINEDQPEQONYDQ-----AV 1169

Qy 183 LGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVV-----VDSGRPIAFGDFEITGTORYP 236
Db 1170 QAA-----NSIINEQTATLDNNAINQAATVNTTKAALHGDVKLQNDKQDHA 1215

Qy 237 EGVSGGLARFOPGTYDLDLLDQQALEQNGHYSGASVQADFRLQGDVRVPVKVSVTEV 296
Db 1216 KQTVSGLAYLNNAGKHMEDTLIDSETT-----RIAVKQDLTEA 1253

Qy 297 K-----RHKLETGI-----RLDSEY---GLGKGIAYDYNNLFPNKGYIGSVVWMDKXET 342
Db 1254 QALDQJLMTLQOOSADKDATRASSAYVNAEPNKKOAYD-----EAVQNAE 1298

Qy 343 TLAAGISOPRNYRGNYWTSNYSYRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFIAE 402
Db 1299 SLIAGLNPNTKNGN--VSSATQAVTYSKN-----GLD---GVERLAQ 1336

Qy 403 GRKTPGSDIDGNSHATMLTASWKRQLNNV 433
Db 1337 DKQTAGNSL-----NHLQDLTPAQQAQALENQI 1363

RESULT 9
US-09-737-149-25
; Sequence 25, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149

; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-737-149-25

Query Match 3.3%; Score 106; DB 10; Length 3034;
Best Local Similarity 20.3%; Pred. No. 4.9;
Matches 129; Conservative 73; Mismatches 207; Indels 228; Gaps 31;

Qy 9 LLPALFFPPHAYAPADLSKENKAAGFALFKNKSP-DTESVKLKPFPVRIDTQDSEIKDM 67
Db 8 VLPALVLLAAALPALEL-----GAAWELURVPGGARAFALPGWSYRLDT----- 53

Qy 68 VEEHLPLITQOEEVLDKEQTG-----GAAWELURVPGGARAFALPGWSYRLDT----- 98
Db 54 -----TRTPRELLDVSRGPAAGRRLLGLGAGTLCGARLAGRLLPLQVRLVARGAPTA 105

Qy 99 VKTMLRSKGYFS-SKVSLTEKDGAYTHITPGPRTKIANGVAILGDI-----SDGNLAE 153
Db 106 PSLVLRARAYGARGVRLRLRSARGAELRSAPVRS-----VPGIGDALCFPAAGGAA- 158

Qy 154 YYRNALENWQPVGSPFDQDSWENSKTSVLCGAVTR---KGYPLAKLGNTRAAV--NPD 207
Db 159 -----SLTSVLEAITNFPACSCPPVAGTCGRGPICLRPG- 193

Qy 208 ATVDLNVVVDGSRPIAFGDFEITGTORYPEQIVSGLARFOPG-TPYDLDLLDQQALEQ 266
Db 194 GSAELRLVLCALGRAAGAVWVELV-----IQATSCPTSPSESPVPSLLNLSOPRAGVVR 247

Qy 267 NGHYSASVQADFRLQGDVRV-----PVKYSVTEVRRHKLETGIRLDSYGLGKGIAYD 320
Db 248 SRRGTGSGTSPQFP-LPSYQVSVPENEPAGTAVTELRAH-----DPDGDAGRLSYQ 298

Qy 321 YYNLFNK---GYI-----GSWV---WDMOKYETTL-----AAGISOPRNYRGNYWTSN 363
Db 299 MEALFEDERSNGYFLIDAATGAVTTARS�DRETQTHVLKVSADVHSGSPRSAAATYLT 358

Qy 364 S-----YNRSTTONLEK-----RAF-----SGGIW--- 383
Db 359 SDTNHSPVFEQSEYRERIRENLEVGYEVLTI RATGDGAPSNANNRYRLLEGAGGVFEID 418

Qy 384 -----YVRDRAGIDARLGAEFIA-----EGRKIPGSDIDGNSHATMLTASWKRQLNNV 433
Db 419 ARSGVVRTRAVVDREAAEAYQLLVEANDQGRN-PPG-----LSAS---ATVHIV 463

Qy 434 LHPENGHYLDGKIGTIGTFLSLSTALIRTSARAGVFFTPENKLGCTFIIRGQA----- 486
Db 464 VEDENDNY-----PQFSEKRYVVO-----VPEDVAVNTAVLRQATDRDQGO 505

Qy 487 ----GYTVARNADVPGLMFRSGGASVRYGYELDSI 519
Db 506 NAAIHYSIVSGNLKQGFYHLHSLSGSLDINPLDFAI 542

RESULT 10

```
US-09-737-149-30
; Sequence 30, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spyttek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15965-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 30
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-737-149-30

Query Match 3.3%; Score 106; DB 10; Length 3034;
Best Local Similarity 20.3%; Pred. No. 4.9;
Matches 129; Conservative 73; Mismatches 207; Indels 228; Gaps 31;

Qy 9 LLPALFPPHAYAPAAOLSENKAAGFALFNKSP-DTESVKLKFPPVRIDTQDSKDM 67
Db 8 VLPALVLLAAALPALEL-----GAAWELRVPGGARAFALGPGWSYRLDT----- 53
Qy 68 VEEHLPLITOOOEVLKDEQGT-----GAAWELRVPGGARAFALGPGWSYRLDT-----FLAEAPDN 98
Db 54 -----TRTPRELLDVSREGPAAGRIGLGCAGTLGCARLAGRLLPLOVLVARGAPTA 105
Qy 99 VKTMLRSKGYFS-SKVSLSITEKDGYTVHTPGPRTKIANVGVAITLGDIL-----SDGNLAE 153
Db 106 PSLVLRARAYGARGCVRLRRSARGAELRSPAVRS-----VPLGDAICFPAAAGGGAA- 158
Qy 154 YRRALENWQOPVGSDFDQDSWENSKTSVLGAVTR-----KGYPLAKIGNTRAAY--NPDT 207
Db 159 -----SLTSVLEAITNFPACSCPPVAGTCRRGPTCLLRPG- 193
Qy 208 ATVDLNVVDVSGRIATGDFEITGTQRYPEQIVSGLARFQPG-TPYDLDDLDDLLDQQALEQ 266
Db 194 GSAELRLVLCALRAGAAWVELV-----IOATSGTSESPSVSPSLNLSQPRAGVVR 247
Qy 267 NGHYSAGSVAQDFRLQGDVRY-----PVKYSVTEVKRHKHLETKGIRLDSYGLGKGIAYD 320
Db 248 SRRGTGSSTSPQFP-LPSYQVSVPENEPAGTAVTELRAH-----DPDEGDAGRLSYQ 298
Qy 321 YYNLENK---GYI-----GSW-----WDMKYEYTL-----AAGISQPRNRYGNWTSNV 363
Db 299 MEALFDRSNGYFLIDAATGAVTTARSUDRETQTHVLKVSADVHGSPPRSRAATYLTVT 358
Qy 364 S-----YNRSTTONLEK-----RAF-----SGIWM--- 383
Db 359 SOTNDHVSFQSEYREIRENLEVGVEVLTIRATDGPAPSNANMRYRLLEGACGVFEID 418
Qy 384 -----YVDRAGIDARLGAFFLA-----EGRKIPGSDIDLGNSHATMLTASWKROLLNNV 433
; Sequence 5703, Application US/09815242

US-09-912-020-299
; Sequence 299, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 299
; LENGTH: 382
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-299

Query Match 3.3%; Score 104.5; DB 10; Length 382;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 79; Conservative 46; Mismatches 142; Indels 87; Gaps 18;

Qy 232 TORPTEQIVSGLARFQGTPTYP-DLDDLDDFOALEQNGHYSGASVQADFRLQGDVVPVK 290
Db 59 TSRSDDQISFGL-----NVPEGDITTSLNYS-----YSNNIWQNDRDHLLAFTLNVP 105
Qy 291 VSVTEVKRHKHLETKGIR-----LDSEYGLGGKIADYDYNLFNKGIVGSVWVWMD-KYEYTL 344
Db 106 FS-----HWMRTDSQAFRNSNASYSMSNDLKGGMNLT--SGVYGTLLPNNLNYSQV 157
Qy 345 -----AAGIS--QPRNRYGNWTSNVSYNRSTTONLEKRAFSGGTWYVDRAGIDAR 394
Db 158 GNTHGGNTSSGTSGYSLNRYGAYGNTNVGYSRSGDSSQIYVYSGGIIAHADGTTFGP 217
Qy 395 LGAEFLAEGRIKPGSDIDLGNSHATMLTASWKROLLNNLVHPENGHYLDGKIGTTLGTFL 454
Db 218 LGDTMVLV--KAPGAD-NVKTENQTHIDWR-----GYAILPFATYERENVALNANSLA 270
Qy 455 SSTALIRT-----SARAGYFTEPKNKLGTFIIRGQAGVTVARDNADVPGLMFR 504
Db 271 DNVELDETIVTVIPTHGAIRATF-----NAQIG-----GKVLMTLKGNSKSVFPGAIVT 320
Qy 505 SGGASSRGVYELDSIGLAGPNSV-----LPERALLVGS-----LEYQLP 544
Db 321 HG-----ENKNGSIVAENGQVYLTGLPQSGQLQVSMGDKNSNCIVEYKLP 366

RESULT 12
US-09-815-242-5703
; Sequence 5703, Application US/09815242
```

Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EUTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5703
; LENGTH: 2025
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; NAME/KEY: VARIANT
; LOCATION: (1)...(2025)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5703

Query Match 3.2%; Score 102; DB 10; Length 2025;
Best Local Similarity 19.4%; Pred. No. 5.8;
Matches 54; Conservative 49; Mismatches 81; Indels 94; Gaps 11;
Qy 176 ENSKTSVLGAVTRKGYPLAKLGNTRAAVNDPDTATVDLNVV-----VDSGRPIAFGDFEI 229
Db 29 QNNKTTI-----KQVQAANSIINEQTATLDNNAIQAATTNTTKAALHGDVKL 77
Qy 230 TGTQRYPEQIVGLARFQPGTPYDLDLLDFFQALQNGHYSGASVQADFRLQQRVPV 289
Db 78 QNDKDHAKQTVSQLAYLNNNAQKHEDTLIDSETT-----RIAV 115
Qy 290 KVSYTEVK-----RHKLETGI-----RLDSEY---GLGGKIADYDYNLFNKGIGSVVW 335
Db 116 KQDTEAQAQLDQMLNTQQSDKADKATASSAYNAEPNKKQAYD----- 160
Qy 336 DMDKYETTAAIGISQPRNRYNYTNSVSNRSTTONLEKRAFSGGIWYVRDRAGIDARL 395
Db 161 EAVQNAESIAGLNNTINKGN--VSSATQAVTTSKN-----GLD--- 198
Qy 396 GAFFLAEGRKTPGSDIDLGNSHATMLTASWKROLNNV 433
Db 199 GVERLAQDKQTAGNSL-----NHLQDLTPAQQAALENQI 232

RESULT 13
US-09-841-132-176
; Sequence 176, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 176
; LENGTH: 982
; TYPE: PRT
; ORGANISM: Chlamydia
; NAME/KEY: VARIANT
; LOCATION: (1)...(982)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-841-132-176

Query Match 3.2%; Score 101; DB 10; Length 982;
Best Local Similarity 21.5%; Pred. No. 2.3;
Matches 133; Conservative 70; Mismatches 237; Indels 178; Gaps 32;
Qy 108 YFSSKVSLEKDGAYTVHITP---GP---RTKIANVGVAI-----LGDILSD 148
Db 335 FFSSNVAAGKGAITYAKKLSVANCQVQFLRNIAIDGGAIYLGSGELSLSDYDGIIFD 394
Qy 149 GNLAETRYRNALENQMQPVGSDFDQDS---WENSKTSVLGAVTRKGYPL-----AKLGN 198
Db 395 GNLK---RTAKENAADVNGVTVSQAISMGGGKITTLRA--KAGHQILFNDEPIEMANGN 449
Qy 199 TRAA-----VNP-----DTATVDLNVVVDSDGR-----PIAFGDFEITGT 232
Db 450 NPAQSSKLLKINDGEGYTGDIVFANGSSTLYQNVTFQGRIVLRERAKLSVNSLSOTGG 509
Qy 233 QRYPE-----QIVSGLARFQPGTPTDLDLLDFFQALE-----QNGH- 269
Db 510 SLYMEAGSTLDFVTPQPPQPPAAANQLITLSNHLSSLSSLLANNAVTPNPPAQDSHP 569
Qy 270 -----YSGASVQADFRLQGRVPVKVSVTEVKRHKLETG-----IRL 307
Db 570 AVIGSTTAGSVTISGPITFFEDLDDTAYDRYDNLGSKNOKINVLKQLQCTKPPANAPSDLT 629
Qy 308 DSE---YGLGG--KIADYDYNLFNKGIGSVVWDMKYEITTLAAGISQPRNRYNYWTSN 362
Db 630 GNEMPKYGYQGSWKLAWDPNTANNPGYTKATWTKTGYNPGPERSVASLVPN---SLMGSI 686
Qy 363 V---SYNRSTTONLEKRAFSGGIW-----YVRDRAGIDARLCAEFLAEGRKIPGSDID 412
Db 687 LDIRSAHSAIQASVDGRSYCRGLMWVSGVSNFFYHDDRDLGO--GYRISGGYSL-GANSY 743
Qy 413 LGNSHATMLTASMKRQLNNVLPENHGYLDGKI-----GTTLGLTFLSSTALIRTSARAG 467
Db 744 FGSMFGLATFEVGRSKDYVVCNSNHHACIGSVYLSLQQALCGSYLFGDAFI---RAS 799
Qy 468 YFFTEPNKKLGTFTIRGQAGYTVARDNADVPISGLMFRSGGASSVGVYELDSIGLAGPNS 527
Db 800 YGFCGNQHK-----TSYTFABE-----SDVR---WNNCLAGEIGA 832
Qy 528 VLP-----ERALLVGSLEYQLPFTRLTSLCAVPHDM---GDAAANPKRMKLKHGS-GLCVR 578
Db 833 GLPIVTPSKLYLNELR---PFVQAEFSYADHESFTBEGDQARAPKSGHLLNLSVVPVCVK 889
Qy 579 W--FSLAP--FSFDIAY 592
Db 890 FDRCSSTHPNKYSEMAAY 907

RESULT 14
US-09-841-132-190
; Sequence 190, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay

```

; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 190
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-190

```

```

Query Watch 3.2%; Score 101; DB 10; Length 1006;
Best Local Similarity 21.5%; Pred. No. 2.3; Mismatches 237; Indels 178; Gaps 32;
Matches 133; Conservative 70;

```

```

QY 108 YFSSKSLTEKDGAYTVHTPP---GP---RTKIANVGVAI-----LGDILSD 148
Db 359 FFSNVAAGGAIYAKKLSVANGCPVQFLRNANDGGAIVLGSGLSLSADYGDIIFD 418
QY 149 GNLAETRYNALENQOPVGSDFODS---WENSKTSVLGAVTRKGYPL-----AKLGN 198
Db 419 GNLK---RTAKENAADVGVTVSSQAISMGGGKITTLRA--RAGHQILFNDPIEMANGN 473
QY 199 TRAA-----VNP-----DTATVDLNVVVDSCR-----PIAFGDFETGT 232
Db 474 NQPAQSKLLKINDGEGYTDIVFANGSSTLYQNVTIEQGRIVLREKAKLSVNSLSQTGG 533
QY 233 QRYPE-----QIVSGLARFOPGPPYDLDDLDFQOALE-----ONGH- 269
Db 534 SLWEAGSTLDFVTPQPPQPPAANQLITLSNLHLSLLANNVNTNPTNPPAODSHP 593
QY 270 -----YGSQVQADFLRLOGDRVPVKVSVTEVKRHKLETG-----IRL 307
Db 594 AVIGSTTAGSVTISGPFIFFEDLDDTAYDRYDWLGSNOKINVLKQLGTGKPPANAPSDLTL 653
QY 308 DSE---YGLGG--KIAIDYNYLNKGYIGSVVWMDKYEITTLAAGISQPNRYGNWTSN 362
Db 654 GNEPKYGYGSKWLADPNTPANNPYTLKATWTKTGYNPGPERVASLVPN---SLWGS 710
QY 363 V---SYNRSTQNLKRAFGGGIW-----YVDRAGIDARLGAELAEGRKIPGSDID 412
Db 711 LDRSAHSAQASVDGSRGRLVWCVSNFFYHDDRDLAQ--GYRIISGYSL-GANSY 767
QY 413 LGNSHATMLTASMKRQLLNVLHPNGHYLDGKI-----GTTLTGTFISSTALIRTSARAG 467
Db 768 FGSMFGLAFTEVFGRSKDYVVCNHHACIGSVYLSQALCGSYLFGDAFI---RAS 823
QY 468 YFTPENKKLGTFIIRQAGTYVARDNADVPGLMFRSGGASSVGRGYELDSIGLAGPNS 527
Db 824 YGFQNOHMK-----TSYTFAE-----SDVR---WDNNGLAGSIGA 856
QY 528 VLP-----ERALLVGSLEVPFTRTLSGAVFHDW---GDAANFKRMKLKHS-GIGVR 578
Db 857 GLPVIPTPSKLYLNELR---PFVQAEFSYADHESFTTEEGDQARAFKSGHLLNLSVPVGK 913
QY 579 W--FSPLAP--FSFDIAY 592
Db 914 FDRCSSTHPNKYSFMAAY 931

```

```

RESULT 15
US-09-815-242-10797
; Sequence 10797, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel

```

```

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10797
; LENGTH: 939
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10797

Query Watch 3.2%; Score 100.5; DB 10; Length 939;
Best Local Similarity 18.3%; Pred. No. 2.3; Mismatches 140; Indels 287; Gaps 36;
Matches 140; Conservative 89;

QY 42 PDRESVK-LPKFPVPRTDQD-----SEIKDMV-----EEHLPLITQ 77
Db 73 PDVSDIGLSPAISIDQKTTSKNPRSTVGTVTEINDYLRLLFARVGHPICPNDHIEITSQ 132
QY 78 QQEVLDKQGTGLAEAPDNVK-----TMLRSKGYFSSKVSLETKDG-----AY 122
Db 133 SVEQMVDK-----VLELPERTKIQLAPVVVVKKGQHKVFMIEQREGVVRMRVDGETY 186
QY 123 TVHTTP-----GPRTKIAN-----VGVAIILGDILSDGNL- 151
Db 187 DVSEAPELEKKNKHDAIIVDRIVVKEGIRSLRFDSEAFALRLAEGVAIVGVIGQEMLF 246
QY 152 AEYI-----RNALENMQQPVGSDFDQSDSWENSKTSVLGAVTRKGYPLAKLGNT 199
Db 247 SEHVACPYCGTGVGELEPRLFSFNAPFGACPCDCG-----LG-----VKLEVD 289
QY 200 RAAVNDP-TATVDLNVVVDGSRPIAFGDFEITGTQRYPEQIVSGLARF--QPCTPYDLDL 256
Db 290 KDLVIPDPTKTLREGAIV-PWNPIIS-----SQYYPQMLEQAATSFGIDMDTTFE-EL 339
QY 257 LLDFQOAL-----EONGHYSASVQADFDRLOGDRVPVKVSVTEVKRHKLETGIRLSEY 311
Db 340 PAQOEIILNGSGEKNFHF---HYENDFGGVROVEVEFEGILKNIKRRIYHETNSDFTRD- 395
QY 312 GLGKTIADYIYNLFNKGYIGSVVWMDKYEITTLAAGISQPNRYGNWTSNVSYNRSTQ 371
Db 396 -----QMRLYWTELTCRSCQ--GYRLNPAQALAVKINGTHIG 429
QY 372 NLEKRAFSGGIWY-----YVRDRAGIDARLGAELAEGRKIPGSDI 411
Db 430 EVSELAIKNAVQPFEGVSVLSEQEITIAIRPILKEVEDRLTFKLVGLDYLTLSR--AAGTL 487
QY 412 DLGNSH-----ATMLTASWKRQLNNVLH-----PENGHYLDG-----KIGTTL- 450
Db 488 SGGEAQRIRLATQIGSN-----LSGVLYILDEPSIGLHQDRNDRLDLSLKKMRDIGNTLI 542
QY 451 -----GTFLSSTALIRTSARAGYF-----FTPENKKGITFIIRGQAGTYVARDNADY 497

```

```
Db 543 VVEHDETMASDYLDIVGPGAGHLGGEIVAAGTPEEVAKNPHSLTQ---YLSCKKVIPV 600
Qy 498 PSGLMFRSGASSVRGYELDSI-----GLAGPNSVLPERALI----- 535
Db 601 PKERRKNGRAIKVTGASENNKNVSEFPPIGEFVAVTGVSGCKSTLVNQILKKALAQK 660
Qy 536 -----VGSLEYOLPFTRLSLGAVFHDMDGDAANFK 565
Db 661 LNRNSNPGKHKSITGYEAIEKIVDIDQSPIGRTPRSPNATYT---SVFDDIRDLFAQTN 717
Qy 566 RMKLKHSGLGVRWFSPLAPSPEDIAVGHSDK-----KIRWH 602
Db 718 EAK-----VRGYKK-GRFSNVKGRCEACKGDGIIKIEMH 752

RESULT 16
US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

Query Match 3.2%; Score 100; DB 10; Length 2086;
Best Local Similarity 19.6%; Pred. No. 9;
Matches 103; Conservative 81; Mismatches 215; Indels 126; Gaps 25;

Qy 80 BEVLDDKQOTGLAEAPDN-VKTMRLSKGYFSSKVSUTE-KDGAYTV-----HITPGPRT 132
Db 967 ENILNKNKSGNLDKAAVENALSQVTNAKGALNGHNHLEQAKSNANTTINGLOHLTTAQKD 1026
Qy 133 KIANVGVAILGDILSDGNLAERYRNALENMQQPVGSDFDQDSDSWENSKTSVLGAVTRKGY 192
Db 1027 KL-----KQVQQAQNVAG-----VDTVKSSANTTINGA----- 1054
Qy 193 LAKLGNTRAAVNPDTAT-----VDLNVVDSGRPIAFGDFEITCTQRYPEQ 238
Db 1055 ---MGTLRNSIQDNTATNNGQNVLDATESNKTNYNNAVDS---ANGVINATSPNNDAN 1107
Qy 239 IVSGLARFQPGTPYDLDLLILDFQALE-QNGHYSGASQVADFRLQDRPVPKVSYTEVK 297
```

```
Db 1108 AINOIAQTQVTSKTNALDGTNHLTOAKQATATNADGAT---NLNKAQKD---ALKAQVTSQA 1162
Qy 298 RHKLETGIRLDSVGLGGKIAYDYNNLFNKGYIGSVVMDMDKYETTLLAAGISQPRNRYGN 357
Db 1163 RVANVTSIQOTA-----NELNTA-MCQLOHGIDDENAT-----KOTQKVRDA 1203
Qy 358 YWTSNVSY-----NRSTQNLEKRAFSGGIWYV---RDRAGIDARLGAEFLEAG 403
Db 1204 EQSKKTAYDQAVAAAKAILNKQTCNSNSDKAAVDRAQLQOVTSTKDALNGDAKL-AEAKAAA 1262
Qy 404 RKTGSDIDILGNSHATMLTASWKROL---LNNVLHPENCHYLDGKIGTTGLTFLSTAL 459
Db 1263 KQNLGLNHTITNAQRTALEGQIQOATTVDGVTNV--KTNANTLDGAMNSLQGSINDKDAT 1320
Qy 460 IRTSARAGYFETTPENKK-----LGTFTIRGOAGYTVARDNADVPGLMFRS----GG 507
Db 1321 LRNO---NYLDADESKRNATQAVTAEGILNKQTCGNTSKADVNDNALNTVTRAKAALNG 1377
Qy 508 ASSVRGYELDSIGLAGPNSVLPPE-RALLVGSLEYQLPFTRTLSG 551
Db 1378 AENLRNKTSAATNI--NG--LPNLTLQKDLKHKHQVEQAQNVAG 1418

RESULT 17
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 3.2%; Score 100; DB 10; Length 5795;
Best Local Similarity 19.6%; Pred. No. 45;
Matches 103; Conservative 81; Mismatches 215; Indels 126; Gaps 25;

Qy 80 BEVLDDKQOTGLAEAPDN-VKTMRLSKGYFSSKVSUTE-KDGAYTV-----HITPGPRT 132
Db 3244 ENILNKNKSGNLDKAAVENALSQVTNAKGALNGHNHLEQAKSNANTTINGLOHLTTAQKD 3303
Qy 133 KIANVGVAILGDILSDGNLAERYRNALENMQQPVGSDFDQDSDSWENSKTSVLGAVTRKGY 192
```

Db 3304 KL-----KQVQQAONVAG-----VDTVKSSANTLNGA----- 3331
Qy 193 LAKLGNTRAAVNPDTAT-----VDLNVVVDGRPIAFGDFEITGTQRYPEQ 238
Db 3332 --NGTLRNSTODTATNNGONYLDATESNKTNTNNAVDS-----ANGVINATSNPNDAN 3384
Qy 239 IVSGLARFQGTGYDLDLLDFOQALE-ONGHYSGASVQADFRLQGDVPPVKVSVTEVK 297
Db 3385 AINOIATQVTSKNALDGTNLTQAKQTATNAIDGAT---NLNKAQKD--ALKAQVTSAQ 3439
Qy 298 RHKLETGRLDSEYGLGKKTAYDYNNLPNKGYIGSVVMDKDKYETTLLAAGISQPRNRYGN 357
Db 3440 RVANVTSIQOTA-----NELNTA-MGOLQHGIDDENAT-----KQTKRYDA 3480
Qy 358 YWTSNVSY-----NRSTTONLEKRAFSGGIIVWV---RDRAGIDARLGAEEFLAEG 403
Db 3481 ESKKTAAYDQAAAKAILNKKQTSNDSKAAVDRALQOVTSTKDALNGDAKL-AEAKAAA 3539
Qy 404 RKIPGSDIDLGNSHATMLTASWKROL---LNNVYHPENGHYLDGKIGTTLTGTLFSLSTAL 459
Db 3540 KONLGLTNHITNAQRTALEGOINQATTVDGVNTV--KTNANTLDGAMNSLQGSINDKDAT 3597
Qy 460 IRTSARAGYFPTPENKK-----LGTFIIRGQAGYTVARDNADVPVSGLMFRS-----GG 507
Db 3598 LRNO---NYLDADESKRNAYTQAVTAAEGLNKKQTGGNTSKADVDNALNTVTRAKAALNG 3654
Qy 508 ASSVRGYELDSIGLAGPNSVLPPE-RALLVGSLEYQLPFTTSLG 551
Db 3655 AENLRNKTSTATNTI--NG--LPMLTQLQKDNLKHQVEQAQNVAG 3695

RESULT 18
US-09-897-056-7
; Sequence 7, Application US/09897056
; Patent No. US200201000741
; GENERAL INFORMATION:
; APPLICANT: JOE, YUJI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: KIDA, TAKASO
; APPLICANT: IGARASHI, DAISUKE
; APPLICANT: OHSUMI, CHIEKO
; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSFORMED PLANT HAVING INCREASED GLUTAMIC
; TITLE OF INVENTION: CONTENT
; FILE REFERENCE: 210425USO
; CURRENT APPLICATION NUMBER: US/09/897,056
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: JP 2000-215279
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-897-056-7

Query Match 3.1%; Score 99.5; DB 10; Length 1025;
Best Local Similarity 19.7%; Pred. No. 3.3;
Matches 113; Conservative 70; Mismatches 225; Indels 167; Gaps 30;

Qy 109 FSSVSLTEKDGAYTHVITPGPRTKIANVGVAILGDLSDGNLAIEYRNALENWQOPVGS 168
Db 41 FHSITLKSKEA-----APVPRP-----VPLSKLTDSPLDGTSVYLEELQRAWEADPNS 91
Qy 169 DFDDSWENSKTSVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNVVVDGRPIAFGDFE 228
Db 92 --VDESMDNFRNFVQ-----ASTSPGISGOTIGESMRLL-----LLVRAYQ 132
Qy 229 ITG-----TORYPEQIVSGLARFQGTGYDLDLLDFOQALEONGHYSVASVQ 276
Db 133 VNGHAKKLDPGLEKREIPDLTPGLYGF---TEADLDREF-FIGVWRMSGFLS----- 183
Qy 277 ADFORLQGDVPPVKVSVTEVRRHLEKLGRLDSEYGLGKKTAYDYNNLKNFNGYIGSVVMD 336

Db 184 -----ENRPV-----QTLRSILS---RLEQAY---CGTIGYEYMHIAADRKCNC---WL 222
Qy 337 MDKYETTLAAGISOPNRYRGNYWTSNYSNRSTTONLEKRAFSGGIYVVRDRAGIDARLG 396
Db 223 RDKLET-----PTPROY---NSERMVVIYDR-LTWSTQFENFLATKWTAKRFGLE---G 270
Qy 397 AEFLAEGRK-----IPGSDIDLGNSHATMLTASWKROLNVLH----- 435
Db 271 AESLIPGKKEMFDSADLGVENIVIGMPHRCRL-----NVLGNVVRKPLQIPESESGG 324
Qy 436 ----PENGHYL-DGKIGTTLTGTLFSLSTALIRTSARAGYFFTPPENKKGTLGTFIIRGQA---G 487
Db 325 TRPDEVGLYGTGDKVYHLGTSYDRPTRGKGKHLHLSLVANPShLEAVDVPVIGKTRAKQ 384
Qy 488 YTVARDNADVPVSGLMFRSGGASSVRYGELDSIGLAG-PN-----GSVLPERALLVG----- 537
Db 385 YTTKDNRTKMGILIHGDSGAFAGVYVETLHLSALPNYCTGGTV-----HIVVNNQVAF 440
Qy 538 -----SLEYQLPFTTSLGAVFHDMDGAAANFKRMKLGSLGSLGVNMFSPPLAFPSFD 589
Db 441 TTDPREGSSQYCTDVAKALSAPIFHVNADDIE-----AVVHACELAAEW---RQTFHSD 492
Qy 590 IA-----YGHSD-----KKIRWHIS 604
Db 493 VVVDLVCYRRFGHNEIDEPSTQPKMKVYKIRSHPS 527

RESULT 19
US-09-747-521-4
; Sequence 4, Application US/09747521
; Patent No. US20020051791A1
; GENERAL INFORMATION:
; APPLICANT: Galloway, Darrel
; APPLICANT: Mateczun, Alfred
; TITLE OF INVENTION: Methods for Protection Against Lethal Infection with Bacillus
; FILE REFERENCE: 22727/04079
; CURRENT APPLICATION NUMBER: US/09/747,521
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-747-521-4

Query Match 3.1%; Score 99; DB 10; Length 764;
Best Local Similarity 18.6%; Pred. No. 2.3;
Matches 111; Conservative 92; Mismatches 221; Indels 172; Gaps 28;

Qy 24 ADLSENKAAGPALPKNKSPTDESVKLKP-----KPPVRIDTQDS-EIKDMVEEHLPLITQ 77
Db 237 SNIHEKK--GLTKYKS-SPEKMWSTASDPYDFEKTGTGRIDKNVSPEARHPLVAAAYPIVHV 293
Qy 78 QOEEVLDKEOTGFLLAEAPDNVKTMLRSKGYFSSKVSUTEKDGAVTVHI--TPGPRTKIAN 136
Db 294 DMENILSKNEDQSTONTDSETRI--SKNTSTSTRTHTSEVHGNAEVANTSTSTHTTSE 351
Qy 137 VGVAILGDILSDGNLAIEYRNALENWQOPVGSDFDDQSDSWENSKTSVLGAVTRKGYPLAKL 196
Db 352 V-----HGN-AEVHAVAIDHSLAG-----ERTWAETMGLNTA---DTARL 389
Qy 197 GNTAAVNPDTATVDLNVVVDGRPIAFGDFEITGTQRYPEQIVSGLARFQGTGYDLDL 256
Db 390 NANIRYVNTGAPI--YNNLPTTS--LVLGKNTLATIKAKENQLS----- 431
Qy 257 LLDFOQALEONGHYSVASV-----QADEFDLQGDVPPVKVSVTEVRRHLEKLGRLD 309
Db 432 -----QILAPNNYPSKNLAPIALNAQDDF-----SSTPTTNNYNQFLELEKTKQLRLDT 481
Qy 310 EYGLGKKTAYDYNN-----LFNKGYIGSV-----VWDMD 338


```
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5330
LENGTH: 801
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5330

Query Match          3.1%; Score 97.5; DB 10; Length 801;
Best Local Similarity 20.1%; Pred. No. 3.3;
Matches 95; Conservative 59; Mismatches 199; Indels 119; Gaps 17;

Qy 25 DLSNKAAGPAL-----FNKSPDTESV-----KLKPPFPVRIDTQDSEIKDMVEEHLPL 74
Db 109 ELRRNSMDAFVPLAYSFSNKENDLKKVTKVLNEQLIPKLQTVGVQNAQLNGQTNREITL 168

Qy 75 ITQOQEEVLDRKEQTCFLAEAPDNVKTMLRSKG-----YFSSKVSILTEKDGAYTVHITPGP 130
Db 169 KFKQNE-----LEKYGLTADDVENYILKTATRTPLGLFQFGDKDKSIIVVDGQY----QSVD 220

Qy 131 RTKIANVGVAIGDILSDGNLAEEYRNALENWQPVGSDFDQDS--WENSKTSVLGAVTRK 189
Db 221 AFKNINIPPLTAG--GOGQ-----SQSDNKNQNSAMSDVNOASPOQNSKASASNNIS-- 271

Qy 190 GYPLAKIGNTRAANVPDTATVDLNVVDSGRPIAFGDEITGTQRYPQIYVSGLARFQPG 249
Db 272 GMPTAKLGD-----LADITY-----GDVRTSISKTNGKDAVN----- 303

Qy 250 TPYDLDDLDFQOALEQNGHYSGASVOADFRLQGDVRPVKVSVEVKRHKLETGIRLDS 309
Db 304 -----LQITKAQDANTVQAKDVQRKIDTFVDENKDFNVTMTDTAKPVEKSL----- 351

Qy 310 EYGLGGKIAYDYNNLFNKGIGSVV-----WDMDKYETTLLAAGISOPRNYRGNYWT--- 360
Db 352 -----YTMVERKASLGTVIAIIVILLFLNRNITTAISIISPLSLLMALIALKL 399

Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 420
Db 400 SDVSLNILT-----YTMVERKASLGTVIAIIVILLFLNRNITTAISIISPLSLLMALIALKL 417

Qy 421 LTKAWKROLLNNVLHPENGVLDGKIGCTTGLTFLSSTALITRSAGYFFTP 472
Db 434 LTDSEEQKLGKGNLIISATTEVFKPIMSTLTIIIVFLPLVFSVSGVGMFRP 485
```

```
RESULT 26
US-09-815-242-12189
Sequence 12189 Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12189
LENGTH: 1055
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12189
```

```
Query Match          3.1%; Score 97.5; DB 10; Length 1055;
Best Local Similarity 20.1%; Pred. No. 5.1;
Matches 95; Conservative 59; Mismatches 199; Indels 119; Gaps 17;

Qy 25 DLSNKAAGPAL-----FNKSPDTESV-----KLKPPFPVRIDTQDSEIKDMVEEHLPL 74
Db 127 ELRRNSMDAFVPLAYSFSNKENDLKKVTKVLNEQLIPKLQTVGVQNAQLNGQTNREITL 186

Qy 75 ITQOQEEVLDRKEQTCFLAEAPDNVKTMLRSKG-----YFSSKVSILTEKDGAYTVHITPGP 130
Db 187 KFKQNE-----LEKYGLTADDVENYILKTATRTPLGLFQFGDKDKSIIVVDGQY----QSVD 238

Qy 131 RTKIANVGVAIGDILSDGNLAEEYRNALENWQPVGSDFDQDS--WENSKTSVLGAVTRK 189
Db 239 AFKNINIPPLTAG--GOGQ-----SQSDNKNQNSAMSDVNOASPOQNSKASASNNIS-- 289

Qy 190 GYPLAKIGNTRAANVPDTATVDLNVVDSGRPIAFGDEITGTQRYPQIYVSGLARFQPG 249
Db 290 GMPTAKLGD-----LADITY-----GDVRTSISKTNGKDAVN----- 321

Qy 250 TPYDLDDLDFQOALEQNGHYSGASVOADFRLQGDVRPVKVSVEVKRHKLETGIRLDS 309
Db 322 -----LQITKAQDANTVQAKDVQRKIDTFVDENKDFNVTMTDTAKPVEKSL----- 369

Qy 310 EYGLGGKIAYDYNNLFNKGIGSVV-----WDMDKYETTLLAAGISOPRNYRGNYWT--- 360
Db 370 -----YTMVERKASLGTVIAIIVILLFLNRNITTAISIISPLSLLMALIALKL 417

Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 420
Db 418 SDVSLNILT-----YTMVERKASLGTVIAIIVILLFLNRNITTAISIISPLSLLMALIALKL 451
```


QY 79 -----QEEVLDRKEQTGFL-----ABEA---PD-NVKTMLR-----104
Db 102 EVNQMIEDRRVDFNQNWYFKLNANSKEAIKPDADVSTWKKLIDLPIYDKWSIFDNDFHESPAQ 161
QY 105 -----SKGYESSKVSUTEKGATVHTTPGPRTKIANVGVAILDLSQNLAEYR 156
Db 162 NEGGQLNGSEAWYRKTFKLDKDLKKNVRLTFD-----GYMDSQVYVNGQLVGHYP 213
QY 157 NALENWQOPVGSDFDDDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVY 216
Db 214 NGYNQFSYDTTKYLOKDGRN--VIAVHAVNKO--PSSRWYSGGIYRVOTLQVTKVHV 269
QY 217 D-SGRPIAFGDFEITGTQRYPEQIYSGLARFQPCPTPYDLDLDDFOALQNGHYSCASV 275
Db 270 EKNGTITLTPKLEEQHGKVEYHTVTSKIVNTD--DKDHELVAEY-QIVERGGHAYTGLV 325
QY 276 QADFDRLQGRVPVKVSVTEVKRHKLET-----GIRLDSYGLGGKI 317
Db 326 RTASRTLKAHESTSLDAILEVERPKLWTVLNDKPALYELITRYVRDQOLVDAKKDLFG-- 383
QY 318 AYDYNYLF-NKGYIGSVVMDKYEITTLAAGIS-----QPRNVYRGNW-----359
Db 384 -YRYHWTNPEGF--SLNGERIKFH-----GVSLHDDHGALGAEEYKAEYRLKONKEM 435
QY 360 ---TSNVYSNRSTQNL-----KRAFSGGIWYVRDRAGIDARLG 396
Db 436 GVNSIRTHNPASEOTLQIAAELGVLLOEAEAFDTWYGGKKPYDYGREFEKDATHPKARKG 495
QY 397 AEFLAERKIPGSDIDL-----GNSHATMLTASWKRQLLNNVLHPNGHYLDCKIGTTL 450
Db 496 EKW-----SDFDLRTMVERGKNPAIF--MWS---IGNEIGEANGD-----AHSI 535
QY 451 GTFLSSPALRTSARAGYFTTPENKKGTLPIRQOAGYTVARDNADVPVSGLMFSGCASS 510
Db 536 ATVKRLVKVTKDVKTRY-----VTMGADKFRFGNG-----SGGHEK 572
QY 511 VRGYELDSIGLAGPNSGVLPERA-----LLVGSLEYQLPFTRTLSGAVPHDMGDAANFK 565
Db 573 IAD-ELDVGFNYSYEDNYKALRAKHPKWLIVGSETTSATRTR--GSYR-----P 619
QY 566 RMKLKHGSG 574
Db 620 ERELKHSNG 628

RESULT 29
US-09-918-909-24
; Sequence 24, Application US/09918909
; Patent No. US2002090704A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Calml, Perry G.
; APPLICANT: Orozco Jr., Emil M.
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: B81166 US CIP
; CURRENT APPLICATION NUMBER: US/09/918,909
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 09/697,367
; PRIOR FILING DATE: 2000-10-26
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 1087
; TYPE: PRT
; ORGANISM: Zea mays
US-09-918-909-24

Query Match 3.1%; Score 97; DB 10; Length 1087;

Best Local Similarity 18.9%; Pred. No. 5.9;
Matches 88; Conservative 63; Mismatches 165; Indels 150; Gaps 19;
QY 16 FPIAYAPAAADLSKNAAGFALFNKSPDTESVKLKPKFP-----VRDQDSEIKD-----66
Db 157 FEKAROYAADLSBLSSEGEKGTNNPESIHDSEMRTPRIGSTDADITWANQHKDKKLY 216
QY 67 --MVEEHLPLIITQQEEVLDRKEQTGFLAE-----EAPD-----97
Db 217 IVLISIHGLRGEMELGRDSDTVKYVVELARALGSGTPGVYRVDLLTRQISAPDVDSYG 276
QY 98 NVMTMLRSKYFSKYSKLSLTKDGAYTVHTTPGPRTKI-----134
Db 277 EPTMELSPISSENFGLGESSGAYIVRIPGPRDKYIPKEHLWPHIQEFVDGALVHMQ 336
QY 135 -----ANVG-----VAILGDLSQNLAEYRYRNALENWQOPVGSDFDQDSHENSKTS 181
Db 337 MSKVLGQIGSGQPVWVPIVTHGYADAGSAALLSGALANVPVFTGHSIGRDK-----389
QY 182 VLGAVTRKGYPLAKLGNTRAAVNP-----DTATVDLNVVYDSGRPI-----222
Db 390 -LDQILKQG-----RQTRDEINATYKIMRRIEAEELCLDTSEI---IITSTROI EQOW 439
QY 223 -AFGDFEITGTQRYPEQIVSGLARF-----QPG-----TPYDLDLDDFOQALE 265
Db 440 GLYDGFDTWARKLRLRIRRGVSCFGRYMPRMIAIPGMEFISHIAPHVDVL-----DSEE 494
QY 266 QNGHYSQA---SVQADFDR-LOGDRVPVKVSVTEVKRHK-LETGIRLDSYEG-----312
Db 495 GNGDGSQSPDPPWADIRFFSNRPKPMILALARPDKKNITITLVKAFGBHRELRNLNL 554
QY 313 ---LGGRIADYNYLFNFKGYIGSVVMDKYEITTLAAGISOPRNYR 355
Db 555 TLINGNRDVIDEMSSSTNAVLTSALKLIDKYD--LYQOVAYPKHHK 598
RESULT 30
US-09-815-242-10728
; Sequence 10728, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10728
; LENGTH: 778
; TYPE: PRT

```
; ORGANISM: Enterococcus faecalis
us-09-815-242-10728

Query Match      3.08; Score 96.5; DB 10; Length 778;
Best Local Similarity 20.48; Pred. No. 3.8;
Matches 115; Conservative 67; Mismatches 166; Indels 217; Gaps 31;

Qy 83 LDKQGTGFLA-----EEAPDNVK-----TMRSGYFSSKVSLSLTKDGAYT 123
   || || || || || || || || || || || || || || || || || || ||
Db 225 LDLPQTALLAGMPQAPNSYDPYTKPDTAKERRDVVLYTM-----YDNKKISAEYERAKA 279

Qy 124 VHIPTGPRTKTYNGVAILGDILSDGNLAEYRNALNWOQPVGSDFDQDQSWENSKTSVL 183
   || || || || || || || || || || || || || || || || || || ||
Db 280 TPIDEG-----LVPLKASDDN-----RKVDVNVYKVEVINEVAKTKGKNVYTDGL 323

Qy 184 GAVTRKGYPLAKLGNTRAAVNPDTATVDLVN-----VVDGSRPIATGCD--FELTGTQRY 235
   || || || || || || || || || || || || || || || || || || ||
Db 324 DIYT-----NLDNNAKQLYDIVNSDQYVAFPPDKMQVAST--- 359

Qy 236 PEQIVSGLARPQPG---TPYDLDL-----LLDFOQALPQNGHYSGAS 274
   || || || || || || || || || || || || || || || || || || ||
Db 360 VIDVASQVRAQIGRRHIPDDVQLGNLAVNTQDVGVSTVKPINDYGPALNLYSTG-- 417

Qy 275 VOAFDRLOGRVPVKVSVTEVKKHLE-----TGIRLDSEYG--- 312
   || || || || || || || || || || || || || || || || || || ||
Db 418 -----RLMVDK-PTKYPGTGTDIDVFNDSLTYQGVITMRRAIMGSRNTTAVQTFDEVGKEN 470

Qy 313 ----LGKGIAYDYNNLFNKGIVGSVWDM--DKY---ETTLAAGISOPRN---YRGNYWT 360
   || || || || || || || || || || || || || || || || || || ||
Db 471 IMPFIKG-LGIDYKNLEASNAISSNTSDVDGDKYGISLSLKUAAYAFANNGIYKPYV 529

Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
   || || || || || || || || || || || || || || || || || || ||
Db 530 NKVVFNDSGTS-----VDYQPDGKRA-----MKDSTAYM 557

Qy 421 LTSWKKQLLNVLHPENG-----HYLD---QKIGTTL-----GTFL 454
   || || || || || || || || || || || || || || || || || || ||
Db 558 MT-----DMLKDVILNGGTGFNGAIPGLIQAKTGTSNYTDEDLARMGTTEKGIAPDSTFV 612

Qy 455 SSTALIRTSARAGY--FFTP---ENKKIGTFPIRGQAGY---TVARNADVPGLMFRSG 506
   || || || || || || || || || || || || || || || || || || ||
Db 613 GYTHYAVSVWGTGNDNRNTPYIQEYGIASDVYREIMSYLSQNVSNDDWVQPDVVV- RVG 671

Qy 507 GASSVR-GYELDSIGLAGPNSVLP 530
   || || || || || || || || || || || || || || || || || || ||
Db 672 NELVVKDAYEV-----PNVQVLP 689

RESULT 31
US-09-815-242-5908
; Sequence 5908, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5908
; LENGTH: 1215
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
us-09-815-242-5908

Query Match      3.08; Score 96.5; DB 10; Length 1215;
Best Local Similarity 19.28; Pred. No. 7.7;
Matches 85; Conservative 66; Mismatches 173; Indels 119; Gaps 19;

Qy 27 SENKAAGFALPKKSPOTESVKLPKFPVRIDT--QDSEIKDMVEEHLPLITQQQ----- 79
   || || || || || || || || || || || || || || || || || || ||
Db 500 SLNTAMGNLI--NATADHQAVEQRCNF--INADTDKQTAYNTAVNEAAAMINKQTCONAQ 556

Qy 80 ---BEVLDKQETGFLAEBAPDNVKTMLRSKGYFSKVSLSLTKDGAYTVHIPTGPRTKI-- 134
   || || || || || || || || || || || || || || || || || || ||
Db 557 TEVEQAITKVQTTLQALNGDHNLOV-----AKTNATQAIIDALT--SLNDPQKLTALKD 606

Qy 135 ---ANVGVAIIGDILSDGNLAEYRNAL-----ENWQOPVGSDF---DQDSEWENSKTSVL 183
   || || || || || || || || || || || || || || || || || || ||
Db 607 QVTAATLVTAHVQIEQNANTLNQAHGLRQSIQDNATKANSKYNEDQPEQNYDQAVQ 666

Qy 184 GAVTRKGYPLAKLGNTRAAVNPDTATVDLVN-----VDSGRPIAFGDFEITGTQRYPE 237
   || || || || || || || || || || || || || || || || || || ||
Db 667 AA-----NNINEQTATLDNNAIQAAATVNTTKAALHGDKVLQNDKDHAK 712

Qy 238 QIVSGLARFQPTPYDLDLLDFOQALPQNGHYSGASVOADFRLQGDVRVPKVSYTEVK 297
   || || || || || || || || || || || || || || || || || || ||
Db 713 QTVSQAHLNNAQKHMEDTLIDSETT-----RTAVKQDLTEAQ 750

Qy 298 RHKLETGIRLDSEY-GLCGKTAYDYNNLFNKGIVGSVWDMDKYE-----TTLAAGISQ 350
   || || || || || || || || || || || || || || || || || || ||
Db 751 -----ALDQLMDALQQSIAADKADTRASSAYNAEPNKKOSYDEAVQNAESIAGLNN 802

Qy 351 PRNYRGNYWTSNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAFLAEGRKIPGSD 410
   || || || || || || || || || || || || || || || || || || ||
Db 803 PTINKNV-----SSATQAV-----ISSKNALD---GVERLAQDKQTAGNS 840

Qy 411 IDLGNSHATMLTASWKROLLNNV 433
   || || || || || || || || || || || || || || || || || || ||
Db 841 L-----NHLDLQLTTPAQQAQALENOI 859

RESULT 32
US-09-815-242-13113
; Sequence 13113, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

Db 488 SYHPDKPGGNRIGLAYSTDGRRNWQYAKERPIVIDNPCKNGDDPSW 53

RESULT 34
US-10-007-693-75
; Sequence 75, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 21021.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 75


```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5779
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5779

Query Match          3.08: Score 95.5; DB 10; Length 841;
Best Local Similarity 17.38; Pred. No. 5.3;
Matches 92; Conservative 77; Mismatches 202; Indels 161; Gaps 22;

QY 92 AEEAPDNVKTMLRSKGYFSSKVSITEKDGAYTVHITPGPRTKIANVGVAILGDIILSDGNL 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 79 SRQLKDNTOATADOPKVTNSDSATVRETSSNQ--SPQNTANQSTTKTSNVTNDKS 135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 AEYRYNALENWQOPVGSDFDQSDNSKTSVGLGAVTRKGYPLAKLGNTRAAVNPDTATVD 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 STTYSN-----ETDKSNLTQA-----KDVSTTPKTTI- 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 LNVVDSGRPIAFGDFEITGTGRTYPEQIVSGLARFPGCTPY-----DLDLLDFQOAL 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 -----KP-----RTLNRMAVNTVAAPQOQGTNVNDKRVHFSNIDIAID----- 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 EQNGHYSGASVQADFRLQGRVPVKVSVTEVKRHKLETGIRLDSEYGLGKIAIYDYNL 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 --KGHVNOTTKTEFWATSSDVLLKANYT-----IDDSVREGDTFTFKYGY 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 FNKGYI-----GSV-----VWMDKYE--TLAAGISOPRYRGNYWTSNVSY 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 FRPGSVRLPSQTONLYNAQGNIAKGIYDSTNTTFTYTNVDQYTNVGRSF--EQVAF 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 NRSTQNLKRAFPSCGIWYVDRAGIDARLCAEFLAEGRKIPGSDIDLGNSHATMLTAS- 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 304 AKRNATTDKTAIK-----MEYTLGNDTYSE-----EIIVDYGNKKAQPLISST 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 -----WKROLLNNVLHPENHYLDGKIGTTLGTFSLSTALIRTSARAGYF----- 469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 348 NYINNEDLSRNWATVNPQNTY-----TKQTFEVLNLGYKFNPAKNKFIYEVTDO 399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 470 -----FTPENKKLGFTFIIRGOAGYTVARDNADVPISGLMFRSGGASSVRGYELDSIGLAG 523
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 400 NOFVDSFTPDTSKLKD--VTQDFDIYSNDNKRTATVDLM--KGQTSNNKQYIIQQV--AY 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 524 PNGSVLPERALLVGSLEYQLFFTRT--LSGAVFHDMDGDAANFKRMKLKHG 572
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 454 PDNSSTDN-----GKIDYTLDTDKYSWSNSYSNVNGSSSTANGDOKKYNLG 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 38
US-09-815-242-12751
; Sequence 12751, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

```

APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR FILING DATE: 2001-03-21
PRIOR FILING DATE: 2000-03-21
PRIOR FILING DATE: 2000-05-23
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-05-26
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-10-23
PRIOR FILING DATE: 2000-11-27
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2000-12-22
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12751
LENGTH: 841
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12751

Query Match 3.0%; Score 95.5; DB 10; Length 841;
Best Local Similarity 17.3%; Pred. No. 5.3;
Matches 92; Conservative 77; Mismatches 202; Indels 161; Gaps 22;

Qy 92 AEEAPDNVKTMLRSGYSSKVSLEKDGAYTHVTPCPRTKIANVGAILDILSDGNL 151
Db 79 SRQKNDTQTATADQPKVTMSDSATVRETSSNMQ---SPQNTANQSTKTSNTVTDNKS 135
Qy 152 AXYRNALENMQPVGSDFDQDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVPDPTATVD 211
Db 136 STTYSN-----ETDKSNLTQA-----KDVSTTPKTTI- 163
Qy 212 LNVVDSGRPIAFGFEITGRTQPEQIVSGLARFQPGTPY-----DLDLLDFQAL 264
Db 164 -----KP-----RTLNRMAVNTVAAPQQTGNVDKVFHFNIDIAID----- 199
Qy 265 EONGHSGASQAADFRLQGRVPKVSVEVKRHKLETGIRLDSYGLGCKIAYDYNNL 324
Db 200 --KHVNQTTCKTEFNATSSDVLKKNY-----IDSVKEGDTFTFKYGOY 245
Qy 325 FNKGYI-----GSV-----VMDMDKYET--TLAAGISQPRNRYGNYWTSNYSY 365
Db 246 FRPGSVRLPSQOTNLYNAQGNIAKGIYDSTNTTFTYFTNVDQYTNVGRSF--EQVAF 303
Qy 366 NRSTONLEKRAFSGGIWYDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTAS- 424
Db 304 AKRNATTDKTAYK-----MEVTLGNDTYSE-----EIIVDYGNNKKAQPLISST 347
Qy 425 -----WKROLLNNLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYF----- 469
Db 348 NYINNEDLSRNTAVYNQPKTY-----TKQTFVNLNLTGYFNPNAKFKIYEVTDQ 399
Qy 470 -----FTPENKKGTLTIIRGQAGYTVARNADVPGLMFRSGGASSVRYGELDSIGLAG 523
Db 400 NOFVDSFTPDTSKLD--VTQDFDIYSNDKNTATVDLM--KGOTSSNKQYIIQOV--AY 453
Qy 524 PNGSVLPERALLVGSLEQLPFTFT-----LSCAVFHDMDGDAANAFKMKLKHG 572
Db 454 PDNSSTDN-----GKIDYTLDTDKYSWSNSYSNVNGSSSTANGDQKYNLG 500

RESULT 39
US-09-897-056-5

Sequence 5, Application US/09897056
Patent No. US20020100074A1
GENERAL INFORMATION:
APPLICANT: JOE. YUJI
APPLICANT: MIWA, TETSUYA
APPLICANT: KIDA, TAKAO
APPLICANT: IGARASHI, DAISUKE
APPLICANT: OHSUMI, CHIEKO
TITLE OF INVENTION: METHOD FOR PRODUCING TRANSFORMED PLANT HAVING INCREASED GLUTAM
FILE REFERENCE: 210425USO
CURRENT APPLICATION NUMBER: US/09/897.056
CURRENT FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: JP 2000-215279
PRIOR FILING DATE: 2000-07-17
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 1017
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-897-056-5

Query Match 3.0%; Score 95.5; DB 10; Length 1017;
Best Local Similarity 19.0%; Pred. No. 7.1;
Matches 104; Conservative 69; Mismatches 200; Indels 175; Gaps 27;

Qy 128 PGPRTKIANVGAILDILSDGNLAEXYRNALENMQPVGSDFDQDSWNSKTSVLGAVT 187
Db 52 PVPRA---VPLSKLTSFLDGTSTVYLELQRAWEADPNS--VDESMDFNFRPVG--- 102
Qy 188 RKGYPKAKIGNTRAANVPDPTATVDLNVVDSGRPIAF-GDPEITGRTQPEQIVSGLARF 246
Db 103 -----QAATSPG---ISQTIQESMRLLLLLVRAQVNG---HMKAKLDPLGLE 144
Qy 247 QGTPYDLDLLDFQQALEQNGHYSGASVQADFDR-----LOGDRVPKVSYTE 295
Db 145 QREIPEDLDLAL-----YGFTEADLDREFPLGVWQMSGFMSNRPV-----Q 186
Qy 296 VKRKLETGIRLDSYGLGCKIAYDYNNLPNKGYTGSVWMDMDKYETTLAGISOPRNYR 355
Db 187 TLRIL---TRLEQAY--CGNIGFEYWHIADRDKN---WLREKLET-----PTPWR 230
Qy 356 GNYWTSNYSYNRSTONLEKRA-----ESGGIWIYVDRAGIDARLGAELAEGRK-- 405
Db 231 -----YNRERREVILDRLANSTOPENFLATKWTTKAKRFGLE---GGESLIPGMKEM 278
Qy 406 -----IPGSDIDLGNSHATMLTASWKROLLNNVLH-----PENGHLYL 442
Db 279 FDRADLGVESIVIGMSHRGL-----NVLGNVVRKPLQIFSEFSGGIRPVDVGYTG 332
Qy 443 DGKIGTTLGT-----FLSSTALIRTSARAGYFFFTPE---NKKL 477
Db 333 TGDVYHLGTSYDRPTRGGKKIHLSLVANPSHLEAADSVVVCKTRAKQYVNDLDRTKNL 392
Qy 478 GTFTIRGQAGYTVARNADVPGLMFRSGGASSVRYGEL-DSIGLAGPNSVLPERALLV 536
Db 393 G-ILIHG-----DGSFAGQGVYETLHLSALPNYTTGCTIHIWVNNQVAFITDPRAG 443
Qy 537 GSLEYQLPFTFTLSCAVFHDMDGDAANAFKMKLKHGSLGVRFSPPLAFPSFEDIA----- 591
Db 444 RSSQYCTDVAKALSAPIFHVNGDDVE-----AVVHACELAAEW---RQTFHSDVVDVLVC 495
Qy 592 ---YGHSD 596
Db 496 YRRFGHNE 503

RESULT 40
US-09-886-468-21
Sequence 21, Application US/09886468
Patent No. US20020037293A1
GENERAL INFORMATION:

APPLICANT: Aventis Pasteur Limited
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
FILE REFERENCE: 77813-5
CURRENT APPLICATION NUMBER: US/09/886,468
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,280
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,281
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,282
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,283
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,284
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,285
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,385
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114,050
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,056
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,057
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,058
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,059
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,061
PRIOR FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 871
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match 3.0%; Score 95; DB 10; Length 871;
Best Local Similarity 19.0%; Pred. No. 6.2; Mismatches 172; Gaps 28;
Matches 128; Conservative 84; Indels 291; Indels 172; Gaps 28;
QY 5 PTALLPALFFPHAYAPA-ADLS-----ENKAAGFALFNKSPDTESVRLKPKFPVVRID 58
DB 6 PLLTSSALVFLPLMAANTDLSDDNYENGSSGSAFTAKETSDAS-----G 54
QY 59 TDSEIKDMVEEHLPLITQOOEEVLDE-----QTGFLABE--APDHVKT 101
DB 55 TTYTLTSDVITNVSAITPADKSCFTNTGGALSFVGADHSLVLQTLIALTHDGAANNNT 114
QY 102 MLRSKGYFSSKVSILEKDGAYTVHITPGRKTANVCVAILGDLSDGNLAERYRNA--- 158
DB 115 ALSFGFSSLLIDSAPATG-----TSGGKAICVTN-----TEGGTATFDNASVT 160
QY 159 -LENWQOPVGSDFQDSWENSKTSVLGAV-----TRKGYPLAKIGNTRAANNPTATVDL 212
DB 161 LQNTSEKGAANVAYSIDILAKTTTALLQNTSTKNGGALCSTANTTVQNSGTVTFSS 220
QY 213 NVVDSGRPIAFGDEITGTORYPEQIVSGLARFPQPTPY-----DLDDLDFQOAL 264
DB 221 NTATDKGGIYSKEKSTLDAN-----TGVVTFKSNKTAKTGCAWSSDDNLALTGNQVL 274
QY 265 EQNCHYSVASQVADFRLQGRDVPVKVSVTEVRHKLKLETGIRLDSEY----- 311
DB 275 FOENKTTGSAQAANNPECGGAIACCYLATATDK-----TGLAISQNOEMSFTSNTTTANG 329
QY 312 -----GLGGKIAYDYNLFNKNKYIGSVVWMDMKYETTLLAA 346
DB 330 GAIYATKCTLDGNTTLTFDQNTATAGCGGAIYETETEDFSKLGSTGTVTFSTNTAKTGGAL 389
QY 347 GISQPRNRYGNYWTSNV--SYNRST---TONLEKRAFSGGIWYVRDRAGIDARLGAEFLA 401

DB 390 YSKGNSSLTGN---TNLLFSGNKATGPNSSANQEGCGGAILAFIDSGSVSDKTGLS-IA 445
QY 402 EGRKIP-----GSDIDLGNSHATMLTASWKQLL-----NNVLHPENGHY-LDGKIG 447
DB 446 NNOEVSILTSNAATVSGGAIYATKCTLTGNGSLTFDGNFAGTSGGAIYTTETEDFTLTGSGT 505
QY 448 TTLGTFLSSTALIRTSARAGYFFTPENKKLGTFIIRGQAGYTVARDNADVPSC----- 500
DB 506 TV--TFSTNTA-----KTGGALYSKGNNSLS-----GNTNLLFSGNKATGPNSSANQEG 553
QY 501 -----LMF-RSGGASSVRGY---ELDSIGLAG-----PNCGSVLPERALLVG--SLEYQLP 544
DB 554 CGGAILSFLESASVSTKKGLWIEDNENVSLSGNTATVSGGAIYATKCALHGNHTTLTPDGN 613
QY 545 FTRTLSCGAVFHDMDG 559
DB 614 TAETAGAIYTTETED 628

Search completed: November 9, 2002, 01:25:07
Job time : 32 secs

Db 228 NNGYAKAQTITKTDVQLNDEKTKVNVITDVGNEGLQYDLSARSARIINLGMGMSAELEPLLSAL 287
Qy 150 NLAEYRNALNWOQPVGSDFDQSDWENSKTSVLGAVTRKCYPLAKLGNTRAAVNPDP-- 206
Db 288 HNDTFRS-----DIADVENAIKAKLG---ERY-----GNTTVNSVPDPPD 327
Qy 207 -TATVDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFPQGPYDLDLDDLLDFOQALE 265
Db 328 ANKTLAITEFVVDAGRRLTVRLREFEGNTVTSADSLRQEMRQOEGTWNSQLVELGKIRLD 387
Qy 266 QNGHYSAGSVAQDFDRLOG--DRVPKVSVEVRKHLETKGIRLDSYGLGCKTAYDYN 323
Db 388 RTGFEE--TVENRIDPINGSNDEVDVYKVKERNTGSINFGYCTESGISYQTSIKODN 445
Qy 324 LFNKGYSVVMQDKYETTLAAGISOPR-----NVRGNVWTSNVSNNRS--TTQNLEKR 376
Db 446 FLGTGAASVSIAGTKNDYGTSVNLGYTEPYTKDGVSLGGNIFPENYDNKSDDTSNNKRT 505
Qy 377 AFSGGI---WYVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG 414
Db 506 TYGNSVTLGPPVNNNSYVGLGHTYKNKISNFALEYNRNLYIQSMKFKGNGIKTNDPFD-- 563
Qy 415 NSHATMLTASWKROLLNNLHPENGHYLDGKIGTTLGTFLLSSTALIRTSARAGYFFTPEN 474
Db 564 -----FSFGWNYNSLARGYPPTKG--VKASLGGRVITPGSDNKYYKLSADVQGFYPLDR 615
Qy 475 KKLCTFIIRGOAGYTVARNADVPGLMFRSGGASSVRGYELDSIGLAGPN----- 525
Db 616 DHRVWSAKASAGYANGFGNKRLLFPYQTYTAGGIGSLRGFAYGSI---GPNAYIAEHNG 672
Qy 526 -----GSLPERALLVGSLEYQLP-----FRT-----L 549
Db 673 TFNKISSDVIGGNATITASAEILVPTPVSDKSTQVTRTSLFVDAASVWNTKWKSDKNGL 732
Qy 550 SGAVFHDMDGDAANFKRMKLGKHSGLGVRFSPFAPSFDA-----YGHSDKKIRWHIS 604
Db 733 ESKVLKOLPDYG---KSSRIRASTGCVGQWOSPICPLVFSYAKPIKKYENDDVE-QFOFS 788
Qy 605 LGTRF 609
Db 789 IGGSF 793

RESULT 2
US-09-135-166-10
; Sequence 10, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jfb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-10

Query Match 6.7%; Score 213.5; DB 3; Length 793;
Best Local Similarity 21.0%; Pred. No. 7.8e-11;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;

Qy 105 SKGYFSKVSILT-----EKDCA-YTVHITPGPRTK-----IANVG--VAILGDILSDG 149
Db 228 NNGYAKAQTITKTDVQLNDEKTKVNVITDVGNEGLQYDLSARSARIINLGMGMSAELEPLLSAL 287
Qy 150 NLAEYRNALNWOQPVGSDFDQSDWENSKTSVLGAVTRKCYPLAKLGNTRAAVNPDP-- 206
Db 288 HNDTFRS-----DIADVENAIKAKLG---ERY-----GNTTVNSVPDPPD 327
Qy 207 -TATVDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFPQGPYDLDLDDLLDFOQALE 265
Db 328 ANKTLAITEFVVDAGRRLTVRLREFEGNTVTSADSLRQEMRQOEGTWNSQLVELGKIRLD 387
Qy 266 QNGHYSAGSVAQDFDRLOG--DRVPKVSVEVRKHLETKGIRLDSYGLGCKTAYDYN 323
Db 388 RTGFEE--TVENRIDPINGSNDEVDVYKVKERNTGSINFGYCTESGISYQTSIKODN 445
Qy 324 LFNKGYSVVMQDKYETTLAAGISOPR-----NVRGNVWTSNVSNNRS--TTQNLEKR 376
Db 446 FLGTGAASVSIAGTKNDYGTSVNLGYTEPYTKDGVSLGGNIFPENYDNKSDDTSNNKRT 505
Qy 377 AFSGGI---WYVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG 414
Db 506 TYGNSVTLGPPVNNNSYVGLGHTYKNKISNFALEYNRNLYIQSMKFKGNGIKTNDPFD-- 563
Qy 415 NSHATMLTASWKROLLNNLHPENGHYLDGKIGTTLGTFLLSSTALIRTSARAGYFFTPEN 474
Db 564 -----FSFGWNYNSLARGYPPTKG--VKASLGGRVITPGSDNKYYKLSADVQGFYPLDR 615
Qy 475 KKLCTFIIRGOAGYTVARNADVPGLMFRSGGASSVRGYELDSIGLAGPN----- 525
Db 616 DHRVWSAKASAGYANGFGNKRLLFPYQTYTAGGIGSLRGFAYGSI---GPNAYIAEHNG 672
Qy 526 -----GSLPERALLVGSLEYQLP-----FRT-----L 549
Db 673 TFNKISSDVIGGNATITASAEILVPTPVSDKSTQVTRTSLFVDAASVWNTKWKSDKNGL 732
Qy 550 SGAVFHDMDGDAANFKRMKLGKHSGLGVRFSPFAPSFDA-----YGHSDKKIRWHIS 604
Db 733 ESKVLKOLPDYG---KSSRIRASTGCVGQWOSPICPLVFSYAKPIKKYENDDVE-QFOFS 788
Qy 605 LGTRF 609
Db 789 IGGSF 793

RESULT 3
US-08-942-046-10
; Sequence 10, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele

```
Query Match      6.7%; Score 213.5; DB 4; Length 793;
Best Local Similarity 21.0%; Pred. No. 7.8e-11;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;
```

QY	105	SKGVSSKVSILT-----EKDGA-YTHITPGPRTK-----IANVG--VAILGDILSDG	149
Db	228	NNGTAKAQITKTDLVOLNDEKTKVNVTIDVNEGLOYDLSARIIGLGMSEALBPLLSAL	287
QY	150	NLAEYIRNALENMQPYGSDFDQDSWENSKTSVLGAVTRKGYPKLAKIGNRAAIVNP---	206
Db	288	HLNDTPRRS-----DIADVENAIAKALG---ERCY-----GNTTVNSVPDPDD	327
QY	207	-TATVDLVNVVDSGRPIAFGDGFETCTQRPEQIVSGLARFPQGPPTPDLDLLDFQQALE	265
Db	328	ANKTLATFTFVDAGRLRTVROLREFEGNTVSADSTLRQMROEGTWNQSJVELGKI RLD	387
QY	266	QNHYSCASQAODPDRLQG--DRYPVKVSVTEVRHKHLEGTGIRDSEYGLGGKTAIDYYN	323
Db	388	RTGFFE--TVENRIDPINGSNDEVVVYKVKERNTGSIFGICYGTESSIYQVISIKODN	445
QY	324	LPNKGYTGSVVDMDKYTETLIAGISOPR-----NTRGNVWTSNVSYNRS--TTQNLEKR	376
Db	446	FLCTGAAVISAGTKNDYGTSVNLGTYEPYTFKDGVSLGGNIFFENIDYDNKSJDTSSNYKRT	505
QY	377	AFSGGI---WYVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG	414
Db	506	TYGSNVTGLGPPVNNESYVGLGHTYNNKISNFALERNLYIQSMKFKNIGIKTNDFD--	563
QY	415	NSHATMLTASKKROLLNNVLHPENGHYLDGKIGTTLGOTFLSSTALIRT SARAGYFFT PEN	474
Db	564	-----PSFGWNYNASLNAGVFYPTKG--VKASLGGGRVITPGSDNKYKYKLSADVOGFYCDR	615

```

Query Match      6.6%; Score 210.5; DB 3; Length 792;
Best Local Similarity 21.5%; Pred. No. 1.5e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps

Qy 134  IANVG--VA1GDI LSGNLA EYRNAL ENWQ P VGSDFDQD SWENSKTSVLGAVTRKGY 191
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270  IGNLGGNSAELEPLLSALHLNDFRRS-----DIADVENAIKAKLG---ERGY 314

Qy 192  PLAKLGNTRA AVNPD-----TATVDLNNVVDSGRPIAFGDFEITGTORYPEQIVGSLARFQ 247
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315  -----GNTTVNSVPDFDDAKNTLAIPTVVDAGRRITVHOLFREGNTVNSADSTLRQEMROO 369

```

```

Query Match      6.6%; Score 210.5; DB 3; Length 792;
Best Local Similarity 21.5%; Pred. No. 1.5e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps

Qy 134  IANVG--VA1GDI LSGNLA EYRNAL ENWQ P VGSDFDQD SWENSKTSVLGAVTRKGY 191
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270  IGNLGGNSAELEPLLSALHLNDFRRS-----DIADVENAIKAKLG---ERGY 314

Qy 192  PLAKLGNTRA AVNPD-----TATVDLNNVVDSGRPIAFGDFEITGTORYPEQIVGSLARFQ 247
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315  -----GNTTVNSVPDFDDAKNTLAIPTVVDAGRRITVHOLFREGNTVNSADSTLRQEMROO 369

```

```
Qy 248 PCTPYDLDLLDFOALEONQHYSCASVQADFDRLQG--DRVPPVKVSVTEVYKRHKLETGI 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 ECTWYNSOLVELGKIRLDRTGFFE--TVENRIDPTNGSDNDEVVYVYKVKERTGTSINFGI 427
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 306 RLDSEYGLGGKIADYNNLFNKGIGSVVWMDKYEITTLAAGISQPR-----NYRGNW 359
:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 GYGTEGSIYQASVKQDNFLGTGAASVITAGTKNDYGTSVNLGYTEPYTKDGVSLGGNVF 487
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 360 TSNVSYNRS--TTONLEKRAFSGGI---WYVRDRAGIDARLG-----AEFLAE----- 402
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNSKSDTSSNYKRTTYGNSVTLGFPVNNENSVYVGLGHTYNNKISNFALEYNRNLYI 547
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 403 -----GRKIPGSDIDLGNSHATMLTASWKRQLLNVLHPENHYLDGKIGITTLGTFLSS 456
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 QSMKFKNGIKTNDFD-----FSGWYNSLNRGYFPYTKG--VKASLGGRTVTPGSD 597
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 457 TALRTSARAGYFFTPENKKGITFIIRGOAGYTVARONADVPGLMFRSGGASSVRGYEL 516
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 NKYYKLSADVQGFYPLDRDHLWVYSASAKAGYANGFGNKRLLPFYQTYTAGGIGSLRGFAY 657
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 517 DSIGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 GSI---GPNALYQGNKFNKISSDVIGCGNAIATASAEELIVPTPFVSDKSONTVRTSLFV 714
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 549 -----LSGAVFHDMDGDAANFKRMKLGKSGLGVRWFSPLAPSFEDIA- 591
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 DAASVWNTKKWSKNGLESNVKLDLPDYG---KSSRTRASTGVGFQWQSPSGPVVFSYAK 771
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 592 -----YGHSDKKIRWHISLGTRF 609
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 PIKKYENDDVE-QFQFSIGGSF 792
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 5
US-09-135-166-8
: Sequence 8, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pete
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MJS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
```

```
TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 792 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-135-166-8
Query Match 6.68; Score 210.5; DB 3; Length 792;
Best Local Similarity 21.58; Pred. No. 1.5e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps 21;
Qy 134 IANYG--VAIIIGDILSDGNLAAYRYNNALENQWQPVGSDFOQDSWENSKTSVLGAVTRKGY 191
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 IGLGGMSAELEPLLALHNDTERRS-----DIADVENAIKAKLG---ERGY 314
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 192 PLAKLGNTRAAVNDP---TATVDLNVVDSGRPIAFGDFEITCTQRYPEQIVSGLARFQ 247
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 -----GNTTVNSVPDFDDANKTLAITFVVDAGRRLLTVIQLRFEGNTVVSADSTLRQEMRQ 369
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 248 PCTPYDLDLLDFOALEONQHYSCASVQADFDRLQG--DRVPPVKVSVTEVYKRHKLETGI 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 ECTWYNSOLVELGKIRLDRTGFFE--TVENRIDPTNGSDNDEVVYVYKVKERTGTSINFGI 427
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 306 RLDSEYGLGGKIADYNNLFNKGIGSVVWMDKYEITTLAAGISQPR-----NYRGNW 359
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 GYGTEGSIYQASVKQDNFLGTGAASVITAGTKNDYGTSVNLGYTEPYTKDGVSLGGNVF 487
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 360 TSNVSYNRS--TTONLEKRAFSGGI---WYVRDRAGIDARLG-----AEFLAE----- 402
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNSKSDTSSNYKRTTYGNSVTLGFPVNNENSVYVGLGHTYNNKISNFALEYNRNLYI 547
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 403 -----GRKIPGSDIDLGNSHATMLTASWKRQLLNVLHPENHYLDGKIGITTLGTFLSS 456
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 QSMKFKNGIKTNDFD-----FSGWYNSLNRGYFPYTKG--VKASLGGRTVTPGSD 597
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 457 TALRTSARAGYFFTPENKKGITFIIRGOAGYTVARONADVPGLMFRSGGASSVRGYEL 516
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 NKYYKLSADVQGFYPLDRDHLWVYSASAKAGYANGFGNKRLLPFYQTYTAGGIGSLRGFAY 657
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 517 DSIGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 GSI---GPNALYQGNKFNKISSDVIGCGNAIATASAEELIVPTPFVSDKSONTVRTSLFV 714
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 549 -----LSGAVFHDMDGDAANFKRMKLGKSGLGVRWFSPLAPSFEDIA- 591
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 DAASVWNTKKWSKNGLESNVKLDLPDYG---KSSRTRASTGVGFQWQSPSGPVVFSYAK 771
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 592 -----YGHSDKKIRWHISLGTRF 609
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 PIKKYENDDVE-QFQFSIGGSF 792
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 6
US-09-942-046-8
: Sequence 8, Application US/08942046
: Patent No. 6264954
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pete
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
```


2IP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-046-8

Query Match 6.6%; Score 210.5; DB 4; Length 792;
Best Local Similarity 21.5%; Pred. No. 1.9e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps 21;

QY 134 IANVG--VAILGDILSDGNLAAYRNALNNQPPVGSDFDQDSDWNSKTSVLGAVTRKY 191
Db 270 IGLNGSAEPLLSALHNDTFRRS-----DIADVENAIAKAGL---ERGY 314
QY 192 PLAKIGNTRAAVNP-----TATVDLNVVDSGRPIAFGDFEITQRYPEQIVSGLARFQ 247
Db 315 -----GNTVNSVPFDDANKTLAITFVDAGRLTVHQLRFEQNTVSADSTLRQMRQ 369
QY 248 PGTPYDLDLDDFOALEQNGHYSGASVQADDFRLOG--DRVVKVSVTEVKRHKLETGI 305
Db 370 EGTWNSQLVGLKIRLDRTEGFE--TVENRIDPINGSNDEVDVYVKVKERTGTSINFGI 427
QY 306 RLDSEYGLGGKIADYDYNLFNKGIGSVWMDKYEYTLAAGISOPR-----NYRGNW 359
Db 428 GYGESGISYQASVKQDNFLGTGAADVSTAGTKNDYGTSVNLGYTEPYFTKDGVSGLGNVF 487
QY 360 TSNSVYRNS--TTQNLKRAFSGGI---WYVRDRAGIDARLQ-----AEFLAE----- 402
Db 488 FENYDNSKSDTSNNYKRTTYSNTVLGPPVNNNSYVYGLHTYNNKISNFALEYNRNLYI 547
QY 403 -----GRKIPGSDIDGNHATMLTASWKQLNNVLPNGHYLDGKIGTTGLTFLSS 456
Db 548 QSMKFKNGIKTDFD-----FSEGWYNSLNRYGFTPKG--VRASLGGRVTIPGSD 597
QY 457 TALIRTSARACYFTPENKKLGTILRQAGYTVARNADVPVSGLMFRSGASSVRGYEL 516
Db 598 NKYYKLSADVQGYPLDRDHLWVYSASAGYANGFGNKRLLPFYQTYTAGGIGSLRGFAY 657
QY 517 DSIGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
Db 658 GSI---GPNALYQONKFNKISSDVIGGNAIATASAEALIVPTFVSDKSONTVTSLFV 714
QY 549 -----LSGAVFHDGMDAAANFKRMLKHGSLGVRWFSPLAFPSFDIA- 591
Db 715 DAASWMTNKKSKNGLESNVLKDLPDYG---KSRTRASTGVCQFQWQSPGQVVPVFSYAK 771
QY 592 -----YGHSDKKIRWHISLGRF 609
Db 772 PIKKYENDDOVE-QPQFSIGGSF 792

RESULT 7
US-08-433-522A-2
Sequence 2, Application US/08433522A
Patent No. 6013514
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-433-522A-2

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHPLITOOQEEVLDEKQOTGFLAEAPDNVKTMLRSKGVFSSKSLT---- 116
Db 184 ESVSSSTLOEOMEYLOPDSWKLGNKPEGAQFEDQLQISIRDYLYNNGYAKAQITKTQVL 243
QY 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAILGDILSDGNLAAYRNALNNQPP 165
Db 244 NDETKVNTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHNDTFRRS----- 296
QY 166 VGSDFDQDSDWNSKTSVLGAVTRKGYPLAKIGNTRAAVNP-----TATVDLNVVDSGRP 221
Db 297 -----DIADVENAIAKAGL---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITQRYPEQIVSGLARFQPTPYDLDLDDFOALBQNGHYSGASVQADDFR 281
Db 344 LTVQLRFEQNTVSADSTLRQMRQEQEGTWNYSQELVGLKIRLDRTEGFE--TVENRIDP 401
QY 282 LQG--DRVVKVSVTEVKRHKLETGIRLDSYGLGGKITAYDYNLFNKGIGSVVWMDMK 339
Db 402 INGSNDEVDVYVKVKERTGTSINFGIGYGTESGYQASVKQDNFLGTGGAASVAGTAGTKND 461
QY 340 YETTLAAGISOPR-----NYRGNWTSVNSYNRS--TTQNLKRAFSGGI---WYVRDR 389
Db 462 YGTSVNLGYTEPYFTKDGVSGLGNVFNENYDNSKSDTSNNYKRTTYSVNLGTPVNNNN 521

REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-433-522A-6

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOBEVLDEKQGTFLAEAPDNVKTMLRSKGYFSSKVSLT----- 116
DB 184 ESVSSTLQOMELQPDMSWMLGNKFEQAFKDLQSIIRDYLLNNGYAKAQITKTDVOL 243
QY 117 --ERDGA-YTVHITPGPRTK-----IANVG--VAIIGDILSDGNLAEEYRNALENWQOP 165
DB 244 NDEKTKNVITDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKIGNTRAAVNPDP-----TATVDLNVVVDGRP 221
DB 297 -----DIADVENAIAKALG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDFQALFQNGHYSGASVOADPDR 281
DB 344 LTVRLRFEQNTVSADSTLRQEMRQOEGTWNLSQVGLKIRLDRDTGFFE--TVENRIDP 401
QY 282 LQO--DRVVPKVSYTEVKRHKLEFGIRLDSYGLGGIAYDYNNLFNKGYIGSVVWMDK 339
DB 402 INGSNDVDVYVKERTNGTSINFGIGYTESGISYQASVKQDNFLGTGAAYTAGTKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTSNVSYNRS--TQNLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSILGGNVFFENYDNSKSDTSNRYKRTYGSNVTGLGPPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROLL 430
DB 522 SYVYGLGHTYKNSINFALEYNRNLYIOSMKFKNGIKTNDP-----FSGWNTNSL 573
QY 431 NNVLHPENGHYLDCKIGTTLTGLTFLSSTALIRTSARAGYFFTPENKKGTLGFIIRGOAGTV 490
DB 574 NRGYFPTKG--VKASLGRVITPGSDNKYKLSADVOGFYPLDRDHLWVWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASSYRGYELDSIGLAGPN-----GSQLPE 531
DB 632 GFGNKRLPFYQTYTAGGIGSLRGFRAYGSI--GPNATVAYGVNGSGTGCTFKKISSDVIGG 688
QY 532 RALLVGSLEYOLP-----FTRLSCAVFDM-----GDAAANFRM----- 567
DB 689 NAIATASAEILVPTFFVSDKSONTVRTSLFVDAASVWNTKWSKDNKLGESDLVLRPLDYG 748
QY 568 ---KLKSGGLGVWSPFLAPESFDIA-----YGHSDKKIRWHISLGTFRF 609
DB 749 KSSRIRASTGVGFOWSPDGLPLVSYAKPIKKYENDDVE--OFQFSIGGSF 797

RESULT 10
US-09-135-166-2
Sequence 2, Application US/09135166
Patent No. 6083743
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,166
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-135-166-2

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOBEVLDEKQGTFLAEAPDNVKTMLRSKGYFSSKVSLT----- 116
DB 184 ESVSSTLQOMELQPDMSWMLGNKFEQAFKDLQSIIRDYLLNNGYAKAQITKTDVOL 243
QY 117 --ERDGA-YTVHITPGPRTK-----IANVG--VAIIGDILSDGNLAEEYRNALENWQOP 165
DB 244 NDEKTKNVITDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKIGNTRAAVNPDP-----TATVDLNVVVDGRP 221
DB 297 -----DIADVENAIAKALG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDFQALFQNGHYSGASVOADPDR 281
DB 344 LTVRLRFEQNTVSADSTLRQEMRQOEGTWNLSQVGLKIRLDRDTGFFE--TVENRIDP 401
QY 282 LQO--DRVVPKVSYTEVKRHKLEFGIRLDSYGLGGIAYDYNNLFNKGYIGSVVWMDK 339
DB 402 INGSNDVDVYVKERTNGTSINFGIGYTESGISYQASVKQDNFLGTGAAYTAGTKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTSNVSYNRS--TQNLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSILGGNVFFENYDNSKSDTSNRYKRTYGSNVTGLGPPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROLL 430
DB 522 SYVYGLGHTYKNSINFALEYNRNLYIOSMKFKNGIKTNDP-----FSGWNTNSL 573
QY 431 NNVLHPENGHYLDCKIGTTLTGLTFLSSTALIRTSARAGYFFTPENKKGTLGFIIRGOAGTV 490
DB 574 NRGYFPTKG--VKASLGRVITPGSDNKYKLSADVOGFYPLDRDHLWVWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASSYRGYELDSIGLAGPN-----GSQLPE 531

```
Db 632 GFGNRLPFYOTYTAGGIGSLRGAYGSI---GPNAIYAEGYNGSGTGTFKKISSDVIGG 688
Qy 532 RALLVGSLEYOLP-----FTRLTSLGAVFHD-----GDAAANFKRM----- 567
Db 689 NAIATASAEILVPTPFVSDKSONVTRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSGGLGVRFSPFLAPSFEDIA-----YGHSDKKIRWHISLGTFRF 609
Db 749 KSSRIRASTGVGFQWSPIGPLVFSYAKPIKKYENDDVE-QQFSGSGSF 797

RESULT 11
US-09-135-166-4
; Sequence 4, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-4

Query Match 6.68; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.08; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

Qy 61 DSEIKDMVEEHLPIITQOOEVLVDKEQTGFLAEAPDNVKTMLRSKGVFSSKVSIT---- 116
Db 184 ESVSSSTLQEQMELQPSWKKLWCKEFGAQFEKDLASIRYYLNNGTAKAQTITDVLQ 243
Qy 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAILGDILSDGNLAERYRNLNMQQP 165
Db 244 NDEKTKVNTIDVNEGLOYLDRSARIIGNLGCMSAEPLLSALHLNDTFRS----- 296
Qy 166 VGSDFDQDSWNSKTSVLGATVTRKGYPLAKLGNTRAAVNPD----TATVDLNVVDSGRP 221
```

```
Db 297 -----DIADVENAIKAKLG---ERYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
Qy 222 IAFGDFEITGQRYPEQIVSGLARFPQCTPYDLDLLDFOALQEQNGHYSGASQVADPDR 281
Db 344 LTVRLRPEGTVTSADSTLRQEMRQOEGTWYNSQLVELGKIRLORTGFFE--TVENRIDP 401
Qy 282 LOG--DRVPKVSVTEVKKRHKLETGIRLDSEYGLGGKIAYDYNNLKNKGYIGSVVWMDMK 339
Db 402 INGSNDEVDVYVKYKERTNGSINFGIGYGTESGISYQASVKQDNFLGTGAASVIASTGKND 461
Qy 340 YETTLAAGISQPR-----NYRGNWTSNVSYNRS--TQNLKRAFPSGGI---WYVRDRA 389
Db 462 YGTSVNLGYTEPYFTKDGVSILGCGNVFFENYDNKSDTSSNYKRTTYGNSVTLGFPVNENN 521
Qy 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNLSHATMLTASWKROLL 430
Db 522 SYVVGGLGHTYNNKISNFALEYNRNLYIOSMKFKGNCIKTNDPD-----FSFGWYNLSL 573
Qy 431 NNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLCTFIIRGQAGYTV 490
Db 574 NRGYFPTKG--VKASLGGRVTIPGSDNKYYKLSADVOGFYPLDRDHLWVWSAKASACAYAN 631
Qy 491 ARDNADVPGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVLP 531
Db 632 GFGNRLPFYOTYTAGGIGSLRGAYGSI---GPNAIYAEGYNGSGTGTFKKISSDVIGG 688
Qy 532 RALLVGSLEYOLP-----FTRLTSLGAVFHD-----GDAAANFKRM----- 567
Db 689 NAIATASAEILVPTPFVSDKSONVTRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSGGLGVRFSPFLAPSFEDIA-----YGHSDKKIRWHISLGTFRF 609
Db 749 KSSRIRASTGVGFQWSPIGPLVFSYAKPIKKYENDDVE-QQFSGSGSF 797
```

```
RESULT 12
US-09-135-166-6
; Sequence 6, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-135-166-6

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOEEVLDKEQTFLAEPADNVKTMRLSKGYFSSKVSILT----- 116
DB 184 ESVSSSTLQEOLEQDPSWMLGNKFGAOFKDLQSIIRDYLLNNGYAKAOKITKTDVOL 243
QY 117 --EKDGA-YTVHITGPRTK-----IANVG--VAILEDILSDGNLAEEYRNALENWQOP 165
DB 244 NDEKTKVNVITDVNEGLQYDLRSARIIGNLGMMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDPDQDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVPD-----TATVDLNVVVDGRP 221
DB 297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFFDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARPOGTPYDLDLLDFOQALEONGHYSAGSVAQADFDR 281
DB 344 LTVROLRPEGNTVSADSTLRQEMROQEGTWYNSQLVELGKIRLDRGTGFFE--TVENRIDP 401
QY 282 LQG--DRVPPVKVSVTEVRKHLETGIRLDSYGLGKAYDYINLNFNGYIGSVVWMDMK 339
DB 402 INGSNDEVDVYKVKERTNGTSINFGIGYGTSGISYQASVKQDNFLGTGAASVAGTNRD 461
QY 340 YETTLAAGISOPR-----NYRGNWTSNVSYNRS--TTQNLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNSKSDTSNNYKRTTYGNSVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPSSDIDLGNHATMLTASWKROL 430
DB 522 SYVVGHLGHTYKINISNFALEYNRLYIQSMKFKNGIKTNDFF-----FSFGWNYSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFLLSSTALIRTSARAGYFFFTPENKKLGTFIIRGOAGTV 490
DB 574 NRGYFTKG--VKASLGRVITPGSDNKYKLSADVQGFYPLDRHLWVWSKASAGYAN 631
QY 491 ARDNADVP SGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531
DB 632 GFGKRLPFQYQYTAGGIGSLRGFPAYGSI--GPNAIYAEGNGSGTGTFFKISSDVIGG 688
QY 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFKRM----- 567
DB 689 NAIATASAEILVPTFPVSDKSONTVRTSLFVDAASVWNTWKSKDNGLESVDVLRPLDYG 748
QY 568 ---KLKHSGLGVRFSPFLAPSFDIA-----YHSDKKIRWHISLGTFR 609
DB 749 KSSRIASTGVGFQWSPIGPLVFSYAKPIKKYENDVE-QPQFSIGGSF 797

RESULT 13
US-08-942-046-2
Sequence 2, Application US/08942046
Patent No. 6264954
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0.. Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942.046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MTS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-046-2

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSEIKDMVEEHLPLITQOOEEVLDKEQTFLAEPADNVKTMRLSKGYFSSKVSILT----- 116
DB 184 ESVSSSTLQEOLEQDPSWMLGNKFGAOFKDLQSIIRDYLLNNGYAKAOKITKTDVOL 243
QY 117 --EKDGA-YTVHITGPRTK-----IANVG--VAILEDILSDGNLAEEYRNALENWQOP 165
DB 244 NDEKTKVNVITDVNEGLQYDLRSARIIGNLGMMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDPDQDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVPD-----TATVDLNVVVDGRP 221
DB 297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFFDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARPOGTPYDLDLLDFOQALEONGHYSAGSVAQADFDR 281
DB 344 LTVROLRPEGNTVSADSTLRQEMROQEGTWYNSQLVELGKIRLDRGTGFFE--TVENRIDP 401
QY 282 LQG--DRVPPVKVSVTEVRKHLETGIRLDSYGLGKAYDYINLNFNGYIGSVVWMDMK 339
DB 402 INGSNDEVDVYKVKERTNGTSINFGIGYGTSGISYQASVKQDNFLGTGAASVAGTNRD 461
QY 340 YETTLAAGISOPR-----NYRGNWTSNVSYNRS--TTQNLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNSKSDTSNNYKRTTYGNSVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPSSDIDLGNHATMLTASWKROL 430
DB 522 SYVVGHLGHTYKINISNFALEYNRLYIQSMKFKNGIKTNDFF-----FSFGWNYSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFLLSSTALIRTSARAGYFFFTPENKKLGTFIIRGOAGTV 490
DB 574 NRGYFTKG--VKASLGRVITPGSDNKYKLSADVQGFYPLDRHLWVWSKASAGYAN 631
QY 491 ARDNADVP SGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531

```

Db 632 GFGNKRPLFPVQTTAGGIGSLRGFAVCGSI---CPNAIYABYGNNGSGTGTFKKLISSDIVIG 688
Qy 532 RALLVGSLEYQLP-----FTRTLSCAVFHDH-----GDAANFKRM----- 567
Db 689 NAIATAELIVPTTFFVSKDSQNTVRTSLFVDAASVWNTKKWSKDKNGLESDVLKRLPDYG 748
Qy 568 ---KLKXGSLGVWRPSPPLAFSEFEDIA-----YHSDKKIRWHISLGTGRF 609
Db 749 KSSRIASTGVGFQWOSPGLPVFSYAKPIKKYKENDDVE-QFQSIGSGSF 797

RESULT 14
US-08-942-046-4
; Sequence 4, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-942-046-4

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps

Qy 61 DSEIKDWVEHLPLITQQQEEVLDKBQTGFLAEAPDNVKTMLRSKGYSKVSILT----- 116
Db 184 ESYSSSTLPQOMELQDPSMMWLMGNKFQAQPEKDLQISRDYLLNNGYAKAKITTKTDVOL 243
Qy 117 --EKDGA-YVYVHTPGPRTK-----IANVG--VALGDLSDGNLAEYRNRALENWQOP 165
Db 244 NDEKTKVNTVIDNVEGLQYDLRSARIIGNLGNMSAELEPILLSAHLNDTFRRS----- 296
Qy 166 VGSDFQDSDWSNKSITSVLGAVTRKGYPLAKLGNTRAAVNPD-----TATYDLNLYVDSGRP 221

```

TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-046-6

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSIKDMVEHLPLITQQOEVLKQGTGFLAEADPNVKTMLRSKGYFSSKVSLSL----116
DB 184 ESYSSSTLQEQMELQDPDSWKLGNKFEQAQFEKDIQSIRDYLLNGYAKAQITKTDVOL 243
QY 117 --EKDGA-YTVHTTPGPRK-----IANVG--VAIGDILSDGNLAERYRNALENWQOP 165
DB 244 NDEKTKVNVITIDNEGLOYDLRSARTIIGNLGSABELEPLLSALHLNDTFRRS-----296
QY 166 VGSDFDQDSHNSKTSVLGAVTRKGYPLAKLGNTRAVNPD-----TATVDLNVVDSGRP 221
DB 297 -----DIADYENAIKAKLG---ERGYGSATVNSV-----PDFDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTORYPQIVSGLARFOPGTPYDLDDLDDFQQALEQNGHYSYGASVQADDFR 281
DB 344 LTVRQLRFEGNTVSASTLRQEMRQEGTWNYSQVELGKIRLORTGFPE--TVENRIDP 401
QY 282 LOG--DRVVPKVSVEVKRHKLETLRDLSEYGLGKIAVDYNNFNKGYIGSVVWMDK 339
DB 402 INGSDEVDVYKVKERNTGTSINFGIGYGTESGISYQASVKQDNELGTGAASVIACTKND 461
QY 340 YETTLAAGISQPR-----NYRGNYSNTSVNRS--TTQNLKRAFSGGI---WYVRDRA 399
DB 462 YGTSVNLGYTEPTFTKDGVSLSGNVFFENYDNSKSDTSSNYKRTYGSNVTLGFPVNNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROL 430
DB 522 SYVGLGHTYKNSIALENRNLYIQSMKFKNGKTKNDFF-----FSGWNYNSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTLFSLTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTV 490
DB 574 NRGVFTPKG--VRASLGRVTPGSDNKYKLSADVQGFPLDRDLHVVWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASVRYELDSICLAGPN-----GSLVPE 531
DB 632 GFGNKRLLPFYQTYTAGGIGSLRGFAYGSI---GPNAIYAEYGNCGSGTGTGFKRISSDVIG 688
QY 532 RALLVGSLEYOLP-----FTRLGSAVPHDM-----GDAANFKRM---567
DB 589 NATATASAEILVTPPVSDKXQNTVTSLEFVDAASVWNTKWKSKDKNGLESVDLKRLPDYG 748
QY 568 ---KLKHGSLGVWFSPLAPFSFDIA-----YCHSDKKIRWHISLGTFR 609
DB 749 KSSRIASTGVGVQWQSPICGPLVFSYAKPIKKYENDDVE-QQFQSIGGSF 797

RESULT 16
US-09-346-408-6
Sequence 6, Application US/09346408B
Patent No. 6338966
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Anderson, Shawn
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
FILE REFERENCE: BB-1167-A
CURRENT APPLICATION NUMBER: US/09/346,408B
EARLIER FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: 60/092,833

EARLIER FILING DATE: July 14, 1998
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Microsoft Office 97
SEQ ID NO 6
LENGTH: 465
TYPE: PRT
ORGANISM: Glycine max
US-09-346-408-6
Query Match 3.9%; Score 123; DB 4; Length 465;
Best Local Similarity 21.6%; Pred. No. 0.0077;
Matches 108; Conservative 67; Mismatches 183; Indels 142; Gaps 28;
QY 98 NVKTMRLRSKGYFSSKVS--LTERKDAYTVHITPG-----PRTKIANYGVAI 141
DB 33 NVFVNFARRHGVGRVSNALIEPDGKGLVELVTVDFERDLKGEALSPLRIKLSRIDL-- 90
QY 142 LGDILSDGNLAERYRNALENWQOPV-----GSDFDQDSHNSKTSVLGAVTRKGYPLA-- 194
DB 91 -----EWVHVLSEGWATPLKGFMRAREFLQTLHFNSLRLLDDGSVNNMSVPIVLA 139
QY 195 -----KLGNTAAVNPDTATVDLNVVDSGRPIA-FGDFEITGTORYP--FOIVSGLA 244
DB 140 IDDAOKHRIGD-----NKKVALFD-----SKGDPVAILNNIEI---YKHPKEERIARTWG 186
QY 245 RFOPGTPY-----DLDLL--LDFQQALEQNGHY--SGASVQADFDRLQGD 285
DB 187 TIAPGLTYVEQITNAGNWLIGDLEVIPIQYNDGLD---HFLSPAQLRAEFTRNAD 243
QY 286 RV-----PV-----KVSYTEVKRHKLETGIR--LDSEYGLGKIAVDYNNFNKGYIGS 332
DB 244 AVFAFQLRNPVHNGHALLMTDTRKLEMGYKPNVLLHPLGG-----YTKADDVP 294
QY 333 VVWDMKYETTLAAGISQPRNYRGNTWTNSVSNRSTTQNLKRAFSGGIWYVRDRA 392
DB 295 LDWRMKQHEKVELDGVLDPEVTTVVSIFPSPMHYAGTEVQ-----WHAKAR--IN 342
QY 393 ARLGAEFLAEGRKIPGSDIDLGNSHAT---MLTASWKROLNNLHPENGHYLDGKIGT 448
DB 343 A--GANFYIVGRD-----PAGNSHPVEKRDLYDADHGKVLKSMAPGLERLNILPRV-- 392
QY 449 TLGTLFSLTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTVARNADVPSGLMFRSGGA 508
DB 393 -----AAYDKTQKMA--FFDPSRPQDPLF--ISGTMKMTLARNKESPPDGFMCPCGWK 442
QY 509 SSVRGYELDSICLAGPNSV 528
DB 443 VLVDYV--DSLVLSS--NGKV 459

RESULT 17
US-09-346-408-4
Sequence 4, Application US/09346408B
Patent No. 6338966
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Anderson, Shawn
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
FILE REFERENCE: BB-1167-A
CURRENT APPLICATION NUMBER: US/09/346,408B
CURRENT FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: 60/092,833
EARLIER FILING DATE: July 14, 1998
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 465
TYPE: PRT
ORGANISM: Momordica charantia
US-09-346-408-4

Query Match		3.8%	Score 121.5;	DB 4;	Length 465;
Best Local Similarity		21.7%;	Pred. No. 0.011;		
Matches 98;		Conservative 63;	Mismatches 177;	Indels 113;	Gaps 24;
Qy	130	PRTKIANVGVAILGDIISDGNLAERYNALENMOQPV-----GSDPDODSWENSKTSVLG	184		
Db	81	PRIELSSIDIQWV-HVLSEG-----WASPLTGFMRSEFLTJHFNLSRLPDG	127		
Qy	185	AVTRKGYPLA-----KLGNTRAAVNPDTATVDLNVVDSGRPIA-FGDFEITGTQRYPEQ-	238		
Db	128	SVANMSVPIVLAIDDAQKRIADSTSVALFDAN-----NNPIALKDIEI-----YKHPEEE	179		
Qy	239	-IVSGLARFQPGTPY-----DLDLL--LDFQQALEQNHYSYGASVQADFD	280		
Db	180	RIARTMTTAPGLPYVDQAITNAGNWLIGDLEVIPIKYHDGLDR-FRQSPAELREEFT	238		
Qy	281	RLOQDRV-----PV-----KVSVTEVKRHKLEGTIR--LDSEYGLGCKTAYDYNNLPNK	327		
Db	239	RRNADAVFAFQLRNPVINGHALMTDTRRLLOMGYKNPILLHLPLGG-----YTK	289		
Qy	328	GYIGSVVMDMKYETTLAAGISOPRNYRGNYWTSNVSYNRSTQNLEKRAFSGGIWTVRD	387		
Db	290	ADDVPLSWRMKQHEKVLGDPETTVSVSIFPSPMHYAGTEVQ-----WHAKA	339		
Qy	388	RAGIDARLGAFFLAEGRKIPGSDIDLGNSHAT-----MLTASWKRQLNNVLPHPENGHYLD	443		
Db	340	R--INA--GANFYTVGRD-----PAGMGHPTEKRDLDYDADHGKVKVLSMAPGLERLNL	389		
Qy	444	GKIGTTLGTLSSPALRTSARAGYFFTPENKKGITFIIRGOAGYTVARNADVPVPSGLMF	503		
Db	390	FRV-----AAYDKTOGKMA-FFDPSRPQDFLF-ISGTMKMTLAKKNPPEGFMC	437		
Qy	504	RSGGASSVRGVYELDSIGLAGPNGSVLPERAL	534		
Db	438	PGGKVKLVVEYD-----SLVPASNDRLPEPVL	464		
RESULT 18					
US-09-346-408-8					
; Sequence 8, Application US/09346408B					
; Patent No. 6338966					
; GENERAL INFORMATION:					
; APPLICANT: Allen, Steve					
; APPLICANT: Anderson, Shawn					
; APPLICANT: Falco, Carl					
; APPLICANT: Rafalski, Antoni					
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins					
; FILE REFERENCE: BB-1167-A					
; CURRENT APPLICATION NUMBER: US/09/346,408B					
; CURRENT FILING DATE: 1999-07-01					
; EARLIER APPLICATION NUMBER: 60/092,833					
; EARLIER FILING DATE: July 14, 1998					
; NUMBER OF SEQ ID NOS: 12					
; SOFTWARE: Microsoft Office 97					
; SEQ ID NO 8					
; LENGTH: 461					
; TYPE: PRT					
; ORGANISM: Triticum aestivum					
US-09-346-408-8					
Query Match		3.6%	Score 114;	DB 4;	Length 461;
Best Local Similarity		22.3%;	Pred. No. 0.051;		
Matches 109;		Conservative 48;	Mismatches 195;	Indels 136;	Gaps 26;
Qy	104	RSKGYSKSVSLTEKDCAYTVHIT--PG-----PRTKIANVGVAILGDIISDGL	149		
Db	38	RRAGMSAIRSLIDPDGGALVDLVAPPGSSRAALRAEAALPRVKLAADV-----	87		
Qy	150	NLAERYNALENMOQPVGS-----DFDQDSWENSKTSVLGAVTRKGYPLA-----KLGNTR	200		
Db	88	---EWAHVLAEGWASPLRGFMREHEYIQLCHFNLSRLPSGGLANWSLPIVLAVDAAKDR	144		
Qy	201	AAVNPOTATVDLNVVDSGRPIAFGDFEITGTQR-----YP-----EQIVSGLARFQPGTPY	252		
US-08-565-386-11					
; Sequence 11, Application US/08565386					
; Patent No. 5741697					
; GENERAL INFORMATION:					
; APPLICANT: Bavoil, Patrik M.					
; APPLICANT: Hsia, Ru-ching					
; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI					
; NUMBER OF SEQUENCES: 23					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP					
; STREET: Clinton Square, P.O. Box 1051					
; CITY: Roches					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; COMPUTER: IBM PC compatible					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: PatentIn Release #1.0, Version #1.30					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/565,386					
; FILING DATE:					
; CLASSIFICATION: 435					
; ATTORNEY/AGENT INFORMATION:					
; NAME: Timian, Susan J.					
; REGISTRATION NUMBER: 34,103					
; REFERENCE/DOCKET NUMBER: 176/60040					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: 716-263-1636					
; TELEFAX: 716-263-1600					
; INFORMATION FOR SEQ ID NO: 11:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 596 amino acids					
; TYPE: amino acid					
; STRANDEDNESS: not relevant					
; TOPOLOGY: linear					
; MOLECULE TYPE: peptide					
US-08-565-386-11					
Query Match		3.6%	Score 113.5;	DB 1;	Length 596;
Best Local Similarity		23.5%;	Pred. No. 0.088;		
Matches 107;		Conservative 55;	Mismatches 214;	Indels 79;	Gaps 22;
Qy	14	FETPHAYAPAADLSENKAAGFALFKKNSPDOTESVKLKPKFPVRIDTODSEIKDWEHLP	73		
Db	118	YFVPQVTSPPGGYAENSIYDFGLPTKVANYRH-QVLPRLAYNLIFWEYRDENIQESLP	176		

; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 29310-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 677-7000
; TELEFAX: (415) 677-7522
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 983 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-164-292B-26

Query Match 3.5%; Score 111; DB 2; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

Qy 3 IKPTALL----LPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLPKPF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL----AVKVGPGLSLDSN 128
Qy 54 -PVRIDTQD----SEIKDMVEEHLPLITQQOEEVLDEKQTFGLAEAPDNVYKMLRSKG 108
Db 129 NALQVHTGDGLTVTDKVSNTQAPLSTTSAGLSLLGLPSLHGLEERLTVTNT----- 181
Qy 109 FSSKVSLETEKDGVTVHTPGPRTKIANVGVAIIGDILSDGNLAEYRYNALENWQOPVGS 168
Db 182 ---GAGLQISNNALAVKVGSGITVDAQNQLAASLGDL----- 216
Qy 169 DFODDSWENSKTSVLG---AVTRKGYPLAKLGNTRAANPDATVTLNVLVVDGSRPIAFG 225
Db 217 ---ESRDNKTVVKAGPGLTITNQALTVA-TGN-GLQVNP-GLQLNITAGQGLNFANN 269
Qy 226 DFETGTQRTPEQIVSGLARFPQPT-----PYD-LDL-----LLDFOQAL 264
Db 270 SLAV-----ELGSGL-HFPFGQNVSLYPCGDIDIRDNRVTVVPAGPGLRMLNHOLAV 320
Qy 265 EQNGHYSASVQADFRLQGDGRVPVKVSVT---EVKRHKLETGIRLDS-----EYGLGG 315
Db 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGTGDDSGRSVVRTGRGL 372
Qy 316 KIAYDYNNLFN-KG-YIGSVVWMDKYETTLAAGISOPRNYRGNWYTSNYSNRSTTONL 373
Db 373 RVANGOVQIFSGRGTAGT-----DSSLTLNIRAPLOFSGPALTASLOGSGPITYNS 424
Qy 374 EKRAF-----SGGIWYVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTFTGLSIGPGMWVDQNRQVNPAGLVFQGNLNVLPNADPLAISDSKISLSLPGGLTQ 484
Qy 417 HATMLTASMKRQLLNVLHPENGHYLDGKIGTTLTGTLFSSTALIRT SARAGYFPTPENKK 476
Db 485 ASNALTLS-----LGNLEFSN-QAVAIAKAGRGL-REFSSQALESSTLVGNGLT----- 532
Qy 477 LGTFIIRQAGYTV-ARDN---ADVPSGLMFRSGCASSVRGYELDSIGLAGPNSGVLP-- 530

Qy 477 LGTFIIRQAGYTV-ARDN---ADVPSGLMFRSGCASSVRGYELDSIGLAGPNSGVLP-- 530
Db 533 LTDVIRPNLGDGLVEVRDNKIIVKLGANLRFENGAVTA-----GTYNPSAPEAPPT 583
Qy 531 ---ERALLVGSLEYQLPFTTSLGAVFHD-----MGDA-AANFKRMKMLKHGSLGVLR 578
Db 584 LTAEPPLRASNSHLQLSLS---EGLVYHNNALALQLGDGMEVNHQGLTLRVGSLQMR 638

RESULT 26
US-08-845-623-26
; Sequence 26, Application US/08845623A
; Patent No. 6001591
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: TIROO, SURESH K.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: BOVINE ADENOVIRUS 3 GENOME
; FILE REFERENCE: 293102002120
; CURRENT APPLICATION NUMBER: US/08/845,623A
; EARLIER FILING DATE: 1997-04-25
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
; US-08-845-623-26

Query Match 3.5%; Score 111; DB 3; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

Qy 3 IKPTALL----LPALFFPHAYAPADLSENKAAGFALFNKSPDTSVKLPKPF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL----AVKVGPGLSLDSN 128
Qy 54 -PVRIDTQD----SEIKDMVEEHLPLITQQOEEVLDEKQTFGLAEAPDNVYKMLRSKG 108
Db 129 NALQVHTGDGLTVTDKVSNTQAPLSTTSAGLSLLGLPSLHGLEERLTVTNT----- 181
Qy 109 FSSKVSLETEKDGVTVHTPGPRTKIANVGVAIIGDILSDGNLAEYRYNALENWQOPVGS 168
Db 182 ---GAGLQISNNALAVKVGSGITVDAQNQLAASLGDL----- 216
Qy 169 DFODDSWENSKTSVLG---AVTRKGYPLAKLGNTRAANPDATVTLNVLVVDGSRPIAFG 225
Db 217 ---ESRDNKTVVKAGPGLTITNQALTVA-TGN-GLQVNP-GLQLNITAGQGLNFANN 269
Qy 226 DFETGTQRTPEQIVSGLARFPQPT-----PYD-LDL-----LLDFOQAL 264
Db 270 SLAV-----ELGSGL-HFPFGQNVSLYPCGDIDIRDNRVTVVPAGPGLRMLNHOLAV 320
Qy 265 EQNGHYSASVQADFRLQGDGRVPVKVSVT---EVKRHKLETGIRLDS-----EYGLGG 315
Db 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGTGDDSGRSVVRTGRGL 372
Qy 316 KIAYDYNNLFN-KG-YIGSVVWMDKYETTLAAGISOPRNYRGNWYTSNYSNRSTTONL 373
Db 373 RVANGOVQIFSGRGTAGT-----DSSLTLNIRAPLOFSGPALTASLOGSGPITYNS 424
Qy 374 EKRAF-----SGGIWYVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTFTGLSIGPGMWVDQNRQVNPAGLVFQGNLNVLPNADPLAISDSKISLSLPGGLTQ 484
Qy 417 HATMLTASMKRQLLNVLHPENGHYLDGKIGTTLTGTLFSSTALIRT SARAGYFPTPENKK 476
Db 485 ASNALTLS-----LGNLEFSN-QAVAIAKAGRGL-REFSSQALESSTLVGNGLT----- 532
Qy 477 LGTFIIRQAGYTV-ARDN---ADVPSGLMFRSGCASSVRGYELDSIGLAGPNSGVLP-- 530

```
Db 533 LDTVLRPNLGDGLEVRDNKIIVKLGANLRFENGAVTA-----GTVPNSAPEAPPT 583
QY 531 ---ERALLVGSLEYQLPFTRTLSCAVFHD-----MGDA-AANFKRMKLGKHSGLGVR 578
Db 584 LTAEPPLRASNSHLQLSL-----EGLVYVHNALALQLDGMEVNOHGLTLRVGSGLOMR 638

RESULT 27
US-08-815-927-26
; Sequence 26, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVIC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
; FILE REFERENCE: 293102002101
; CURRENT APPLICATION NUMBER: US/08/815,927
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
US-08-815-927-26
```

Query Match 3.5%; Score 111; DB 3; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

```
QY 3 IKPTALL-----LPALFFPHAYAPADLSKKAAGFALFKNKSPTDESVKLKPKF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL-----AVKVGPLSLDSN 128
QY 54 -PVRIDTQD-----SEIKDMVEEHLPLITQOOEEVLQKEQTGFLEAEAPDNVKTMLRSKY 108
Db 129 NALQVHTGDGLTVTDKVSLSNTQAPLSTTSAGLSLLGLPSLHGLGEERLTNT----- 181
QY 109 FSSKVSLTEKDGAYTHITPGPRTKIANVGVAILGDILSDGNLAEYRNALENWQOPVGS 168
Db 182 ---GAGLQTSNNALAVKVGSGITVDAQNLAAASLGDL----- 216
QY 169 DFDQDSWENSKTSVLG---AVTRKGYPLAKGNTRAAVNPDTATVDLNVVVDVSGRPIAFG 225
Db 217 ---ESRDNKTVYKAGPGLTITNQALTV--TGN-GLOVNP--GOLQLNITAGOGLNFPANN 269
QY 226 DFEITGTORYPEQIVSGLARFPQGT-----PYD-LDL-----LLDFOOAL 264
Db 270 SLAV-----ELGSGL-HFPFGQNVSLYPGDGIDIRDNRVTVVPAGPGLRMLNHQLAV 320
QY 265 EQNGHYSVASQVADFRLQDGRVPVKVSVT---EVRKHLETGIRLDS-----EYGLGG 315
Db 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGIGTDDSGRSVVRTGRGL 372
QY 316 KIAIDYNNLFN-KG-YIGSVVWMDKYETTTLAAGISOPNRYNRYNWTSNVSYNRSTTONL 373
Db 373 RVANGQVQIFSGRGTAGT-----DSSLTLNIRAPLQFSGPALTASLOGSGPITYNS 424
QY 374 EKRAF-----SGGIWYVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTFGLSIGPGMWVDQNRLOVNPAGLVFQGNLNVLPNADPLAISDSKISLSILGPGLTQ 484
QY 417 HATMLTASWKROLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKK 476
Db 485 ASNALTL-----LGNGLEFSN--QAVAIKAGRGL-RFESSQALESLSLTGVNGLT----- 532
QY 477 LGTFIIRGQAGYTV-ARDN---ADVPSGLMFRSGGASVVRGYELDSITGLAGNVSVP-- 530
```

```
Db 533 LDTVLRPNLGDGLEVRDNKIIVKLGANLRFENGAVTA-----GTVPNSAPEAPPT 583
QY 531 ---ERALLVGSLEYQLPFTRTLSCAVFHD-----MGDA-AANFKRMKLGKHSGLGVR 578
Db 584 LTAEPPLRASNSHLQLSL-----EGLVYVHNALALQLDGMEVNOHGLTLRVGSGLOMR 638

RESULT 28
US-09-103-330-26
; Sequence 26, Application US/09103330A
; Patent No. 6319716
; GENERAL INFORMATION:
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: ISOLATION OF MUTANTS IN THE E3 REGION OF THE
; FILE REFERENCE: 293102002121
; CURRENT APPLICATION NUMBER: US/09/103,330A
; CURRENT FILING DATE: 1998-06-23
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: 08/164,292
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
US-09-103-330-26
```

Query Match 3.5%; Score 111; DB 4; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

```
QY 3 IKPTALL-----LPALFFPHAYAPADLSKKAAGFALFKNKSPTDESVKLKPKF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL-----AVKVGPLSLDSN 128
QY 54 -PVRIDTQD-----SEIKDMVEEHLPLITQOOEEVLQKEQTGFLEAEAPDNVKTMLRSKY 108
Db 129 NALQVHTGDGLTVTDKVSLSNTQAPLSTTSAGLSLLGLPSLHGLGEERLTNT----- 181
QY 109 FSSKVSLTEKDGAYTHITPGPRTKIANVGVAILGDILSDGNLAEYRNALENWQOPVGS 168
Db 182 ---GAGLQTSNNALAVKVGSGITVDAQNLAAASLGDL----- 216
QY 169 DFDQDSWENSKTSVLG---AVTRKGYPLAKGNTRAAVNPDTATVDLNVVVDVSGRPIAFG 225
Db 217 ---ESRDNKTVYKAGPGLTITNQALTV--TGN-GLOVNP--GOLQLNITAGOGLNFPANN 269
QY 226 DFEITGTORYPEQIVSGLARFPQGT-----PYD-LDL-----LLDFOOAL 264
Db 270 SLAV-----ELGSGL-HFPFGQNVSLYPGDGIDIRDNRVTVVPAGPGLRMLNHQLAV 320
QY 265 EQNGHYSVASQVADFRLQDGRVPVKVSVT---EVRKHLETGIRLDS-----EYGLGG 315
Db 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGIGTDDSGRSVVRTGRGL 372
QY 316 KIAIDYNNLFN-KG-YIGSVVWMDKYETTTLAAGISOPNRYNRYNWTSNVSYNRSTTONL 373
Db 373 RVANGQVQIFSGRGTAGT-----DSSLTLNIRAPLQFSGPALTASLOGSGPITYNS 424
QY 374 EKRAF-----SGGIWYVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTFGLSIGPGMWVDQNRLOVNPAGLVFQGNLNVLPNADPLAISDSKISLSILGPGLTQ 484
QY 417 HATMLTASWKROLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKK 476
Db 485 ASNALTL-----LGNGLEFSN--QAVAIKAGRGL-RFESSQALESLSLTGVNGLT----- 532
```

Qy	477	LCFTFLTRGQAGYTV-ARDN----	ADVP	SGLMFRSGGASSYRGVELDSIGLAGP	NGSVLP--	530
			:		:	
Db	533	LTDTVRPNLGDLEVRDNKIIYVKIGANLRF	ENGAVTA-----	GTVNP	SAPEAPPT	583
			:		:	
Qy	531	---ERALLVGSLEVQLPFFTRT	LSGAVFHD-----	MGDA-	AANFRKMKLKHGSGLGVR	578
			:		:	
Db	584	LTAEPPLRASNSHLQSLSL-	---EGLVVHNNALALQLGDGMEVNHQHGL	TLRVCSGLQMR	638	
			:		:	

RESULT 29
 US-09-435-242-26
 ; Sequence 26, Application US/09435242
 ; Patent No. 6379944
 ; GENERAL INFORMATION:
 ; APPLICANT: MITTAL, SURESH K.
 ; APPLICANT: GRAHAM, FRANK L.
 ; APPLICANT: PREVIC, LUDWIK
 ; APPLICANT: BABIUK, LORNA A.
 ; TITLE OF INVENTION: MAMMALIAN CELL LINES EXPRESSING BOVINE ADENOVIRUS
 ; FILE REFERENCE: 293102002102
 ; CURRENT APPLICATION NUMBER: US/09/435,242
 ; CURRENT FILING DATE: 1999-11-05
 ; EARLIER APPLICATION NUMBER: 09/815,927
 ; EARLIER FILING DATE: 1997-03-13
 ; EARLIER APPLICATION NUMBER: 08/164,294
 ; EARLIER FILING DATE: 1993-12-09
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 983
 ; TYPE: PRT
 ; ORGANISM: Bovine adenovirus type 3
 US-09-435-242-26

```

Qy 477 LGFTIIRGQACYTV-ARDN---ADVPSGLMFRSGGASSVRGVEYELDSIGLAGPNGSVLP-- 530
      |         |         |         |         |         |         |         |         |         |
Db 533 LTTDVIRPNLGDGLVEQRDKIIIVKLGANLRFENGAVTA-----GTVNFSAPEAPPT 583
      |         |         |         |         |         |         |         |         |         |
Qy 531 ---ERALLVGSLEVQLPFTRTLSCGVFHD-----MGDA-AANFKRMKCLKHSGSLGVR 578
      |         |         |         |         |         |         |         |         |         |
Db 584 LTAEPPLRASNSHLQLSLS---EGLVVHNNALALQLGDGMVQHNGHGLTRVSGSLQMR 638
      |         |         |         |         |         |         |         |         |         |

RESULT 30
US-09-206-942-67
; Sequence 67, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 67
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-67

```

QY 532 RALLVGS--LEYQLPFT-----RTLSGAVFHDGM-----DAAANFKRMKLKHGSG 574
Db 860 --TLIGSDFNHOKPLTIKRDVINSNLTAGGNIVNAGNLATVESNANEK----- 908
QY 575 LGVRWFSPAPFSDIAYGSHDKIRWHISL---GTRF 609
Db 909 -----AITNFTFNVG-GLFDKNKGSNISIAKGGARF 938
RESULT 31
US-09-376-330-18
; Sequence 18, Application US/09376330
; Patent No. 6399321
; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David F.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGGT
; FILE REFERENCE: 2139-9*US*
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: S. cerevisiae KRE5
US-09-376-330-18
Query Match 3.4%; Score 109; DB 4; Length 1365;
Best Local Similarity 20.7%; Pred. No. 0.92;
Matches 124; Conservative 93; Mismatches 212; Indels 170; Gaps 33;
QY 41 SPOTESVKLPKPPVRDITDQ-SEIKDMVEHLPLITQOOEEVLADKEOTGFLEAEAPDNV 99
Db 208 SLOGKSVEYPLTHPLETLONGSRMSSIPQKKILYVTPKEILVAGDNDQDLHLEPEEL 267
QY 100 KTLRSKGYFSSKVSLETKDGAYTHITPG-----PRTKIANVGVAIIGDILS-- 147
Db 268 RELDLRVTLSISEFYQYKKDITATLNTKTSIVNFPILSKOLIKVSVN---KDIITSN 323
QY 148 -----DGNLAEYRNALENKQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKGNT 199
Db 324 EELNSKGFYDNLGLYING-QNMKITSPTY-----NLITALKTEYQSLLKITNL 372
QY 200 RAAVNPDTATVDLNVVDSGRPIAFGDEITGTQRYPEQIVSGLARFQPGTPPYDLDLLD 259
Db 373 LOLEPSKCILDSKFLUNK-----FSQFSL-----GKLNQOP-IKMDLHTIPG 415
QY 260 FOQA-----LEONGHYSGA--SVQADFRLQGDVRVPKVSVEVKRH-----K 300
Db 416 FSESIVFNDIESDPQYDELVNSVQAFDKSKFGELP-----EIKQWSEIIFVIDFAR 469
QY 301 LETGIRLDSEY--GLGKIAIDYNNFNKGY---IGSVWMDMKYETTLAAGISQPNRYR 355
Db 470 LE-----DSEYKEALGLVR--AVNVYSQGPQRVGLLFPSSSDSKSV----- 511
QY 356 GNYWTSNVSYN-RSTTONL-EKRAFSGGIWVRDRAGIDARLGAETLAEGRK-----IPGS 409
Db 512 -----NKIYELKNSTONLTELKSF-----LETMLADGLSANAKHSHKHIPVP 553
QY 410 DI-----DLGNSHATMLTASMKROLLNNVLP--ENGHYLDGKIGTTLTGLTFLSSTALIRT 462
Db 554 DVPHLLDELQIDETSI-----IINGEYIPFPKNNYLIKV-----IKKDETEIRK 599
QY 463 SARAGYFFTPENKLGTFITRGOAGYTV--RDNADVPVSGLMFRSGGASSVRGYELDSIG 520
Db 600 ELSNS-----SPKNQIS---VRDLLHYKRSANLRHNKYTPN--YFADSVSYSSVNNTALE--- 648

QY 521 LAGPNSGVLPER-ALLVGSLEYQLPFTRTLSGAVFHDGMGAAANFKRMKLKHGSLGVR 578
Db 649 -----SVCSERICYYTKNEEYNLLHTITL-----VDFGSHALKRLNLLHTTSFVGVR 697
RESULT 32
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
Query Match 3.4%; Score 108.5; DB 4; Length 1375;
Best Local Similarity 19.7%; Pred. No. 1;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;
QY 30 KAAGFALFKNKSPDTSVKLPKPPVRIDTODSEIKDMVEEH-----LPLITQOOEEV-- 82
Db 729 RTSQVAVIEGNPS-----LRLKASDRVVVNMGAHKQAVRPLLLTTDNGIKA 777
QY 83 --LDKEOTGEL--ABEAPDNVKTMLRSKGYFSSKVS-----LTKDGAYTHIIT 127
Db 778 YHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSYLVGVVPGVGAADODVRAAATA 837
QY 128 PGPRTKIANVGVAIIGDILSDG-----NLAEYRNA-----LENWQOPVGSDFDQDS 174
Db 838 PSTDGKSVHQNAALDSRVMEFGFSNFQAFATKKKEEYNVVIKKNVDFKFAEAGVTFDEMAP 897
QY 175 WENSKT--SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFG----- 225
Db 898 QYVSSTDGSLDSVIQNGYAF-----TDYDLGI-----SKPNKYGTADDLVK 940
QY 226 -----DPEITGTQRYPE-QIVSGLARFQPGTP-----YDLDLLDF 260
Db 941 AIKALHSGIKVMADWPDQMYALPEKEVVTATRVDRKYGTVPVAGSQIKNTLYVVD--GK 997
QY 261 QOALEQNGHYSGA---SVQADFRLQGDVRVPKVSVEVKRHKLETGLRLDSEYGLGKI 317
Db 998 SSGKDOQAKYGGAFLLELQAKYPEL-----FARKQISTGVMPDPSVKIKOWS 1044
QY 318 A--YDYYNLFNKGIVGSVWDM--DKYETTLLAAGISOPRN--YRGNYYTNS----- 362
Db 1045 AKYENGNTILGRG--AGYVLKQDQATNTYFSLVSDNTLTKPKSLVNPNGHGTSSSVTGLVPDGK 1103
QY 363 --VSYNRSTONLEKRAF--SGGIWYVRDR-----AGIDARLGAE--FLAEGRKIPGDI 411
Db 1104 GYVYVY--STSGNOAKNAFISLGNWYFDNNGYVMVTGAQISNGANYFPLSNGIQLRNAIY 1161

QY 412 DLGNSHATMLTASMKRQLLNVLHPENGHYLDGKIGTTLTGLTSLSTALIRTSARAGYFTT 471
Db 1162 DNGNKVLSYYGNDGRRY-----ENGYYLFGQWRVYFQNGIMAVGLTRVHGAVQVYFDA 1213
QY 472 PENKKLGTFTIIR--GOAGYTVARNADVPSCLMFRSGASSVRCGYELDSIGLA----- 522
Db 1214 SGFOAKGOFITTAQDKLRY-FDRSGNQISNRFRVN---SKGEWFLFDHNGVAVTGTVT 1269
QY 523 -----GPGSVLPERALV---GSLEYOLP-----FTRTLSGAVF-HDMGDAA 561
Db 1270 NGORLYEKPNG--VOAKGEFIRDANGYLRYDPSNGNEVRNFRVNRKSGEWFLEHDHNGIA 1327
QY 562 ANFRMKLKHGSGGLGVRF 580
Db 1328 VTGARVVNGHASILSLMVF 1346
RESULT 33
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 3.4%; Score 108.5; DB 4; Length 1375;
Best Local Similarity 19.7%; Pred. No. 1;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;
QY 30 KAAGFALFKNKPOTESVKLKPFPFRIDTQDSEIKDWVEH-----LPLITQOQEEV-- 82
Db 729 RTSQVAVIEGNPS-----LRLKASDRVVVNGAAHKNQOARPLLLTTDNGIKA 777
QY 83 --LQKEOTGEL--AEEAPDNVKTMLRSKGYFSSKVS-----LTEKDGAYTVHIT 127
Db 778 YHSQEAAGLVRYNDRGELIFTAADIKGYANPOVSGYLGWVPGVGAADODVRVAASTA 837
QY 128 GPRRTKIANGVATLGLDILSDG-----NLAEYRNA-----LENMQQPVGSDFOODS 174
Db 838 PSTDCKSVHQAALDSRVMPGEGSFNFQAFATKKEEYTNVVIKAVNDKFAENGVTDFEMAP 897
QY 175 WENSKT--SVLGATVRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPTAFG----- 225
Db 898 QYVSTSDGSELDVSIQNGYAF-----TDYDLGI-----SKPNKYGTADDLVK 940
QY 226 -----DFETGTQRYPE-QIVSGLARFQGTB-----YDLDDLDF 260
Db 941 AIKALHSGIKVMADWPDQWYALPEKEVTVATRVDRKYGTPVAGSQIKNTLYYVD---GK 997

QY 261 QQALEQNGHYSGA---SVQADFDRLQGRVPRVKYSVTEVKKRHKLETGIRLDSSEYGLGCKI 317
Db 998 SSGKDDQAKYGGAGFLLEELQAKYPYL-----FARKQISTGVPMPSPVKIKQWS 1044
QY 318 A--DYTNLNMKGIGSVVWDM--DKYETTLLAAGISOPRN-YRGNYYTNSN----- 362
Db 1045 AKYFNGTNIILGRG-AGVYLKDOATNTYFSLVSDNTFLPKSLVNPNHGTSSTSVTGLVFDGK 1103
QY 363 --VSYNSTTQNLKRAAF--SGGIWYVRDR-----AGIDARLGAEE--FLAEGRKIPGSDI 411
Db 1104 GYVY--STSGNOAKNAFISLGNWYFDDNNGYVMTCAQSGINGANYFSLNGICQLRWAIY 1161
QY 412 DLGNSHATMLTASMKRQLLNVLHPENGHYLDGKIGTTLTGLTSLSTALIRTSARAGYFTT 471
Db 1162 DNGNKVLSYYGNDGRRY-----ENGYYLFGQWRVYFQNGIMAVGLTRVHGAVQVYFDA 1213
QY 472 PENKKLGTFTIIR--GOAGYTVARNADVPSCLMFRSGASSVRCGYELDSIGLA----- 522
Db 1214 SGFOAKGOFITTAQDKLRY-FDRSGNQISNRFRVN---SKGEWFLFDHNGVAVTGTVT 1269
QY 523 -----GPGSVLPERALV---GSLEYOLP-----FTRTLSGAVF-HDMGDAA 561
Db 1270 NGORLYEKPNG--VOAKGEFIRDANGYLRYDPSNGNEVRNFRVNRKSGEWFLEHDHNGIA 1327
QY 562 ANFRMKLKHGSGGLGVRF 580
Db 1328 VTGARVVNGHASILSLMVF 1346
RESULT 34
US-09-118-319-6
; Sequence 6, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118,319
; CURRENT FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-118-319-6

Query Match 3.4%; Score 108; DB 3; Length 459;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 90; Conservative 63; Mismatches 193; Indels 94; Gaps 19;
QY 141 ILGDILSDGNGIAEYRNALENWQPVGSD-----PDQSWENSK 179
Db 6 VLATLFATGALASEC-----HWQYPCCKDCTVYVYTDTEGKWGLVNDWCMIDNRCSSNN 60
QY 180 TSVLGATVRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQI 239
Db 61 NNCSSITSGQYPCCSNNCKVEYTDN----DGKWGVNNWNCISNCSGGGQOQOQPTQ 116
QY 240 VSGLARFQPTPYDLDLDDLLDPOQALEONGHYSGASVQADFDRLQCD-----RVPVK 290
Db 117 TQPTQPOQTPSSDNF--FENEIYSNKYPQG-EVDISIKKNGDLKAKAEKVKYVPTA 172
QY 291 V-----SVTEVKKRHKLETGIR-----LQSEYGLGCKIADYDYNLNMKGYIG 331
Db 173 VWLAWDGAPOEVPYRLOEAGNKTVFVFLYMIPTRDCCGANASAGSATIDKY----KGYN 228
QY 332 SVVWDMKYEFTLAAGISOPRNRYGNTVSNVSRSTTONLEKRAFSGGIWYVRDRAGI 391
Db 229 NIYNTSNQYKNSKIVMLEPDTI--GNLVTNN--NDNCNRVNRNMHKAQALYSKFCGTOSH 286


```
QY 392 DARLGAEFLAEGRKIPGSDIDLGNSHATMLTASWKROLLNNVLPENGHYLDGKIGTTLG 451
Db 287 KVLDA---AHGAWL-----NOYADQ-TANVIKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKGTLFIIRG-QAGYTVARDNADVPGLMFRSGGASS 510
Db 326 TNVSNYOSIESEYK---YHQNLRALSKGVRGLKFIVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLQORPKGNPNPGSMP 401

RESULT 35
US-09-286-691-2
; Sequence 2, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; CURRENT FILING DATE: 1999-04-05
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-286-691-2

Query Match 3.4%; Score 108; DB 4; Length 459;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 90; Conservative 63; Mismatches 193; Indels 94; Gaps 19;

QY 141 ILGDILSDGNLAEYRNALENWQOPVGS-----FDQDSWNSK 179
Db 6 VLATLFATGALASEC-----HWQYPCCKDCTVYVYTDTEGKWGLNNDWCMIDNRCSSNN 60
QY 180 TSVLGAVTRKGYPLAKIGNTRAAYNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQI 239
Db 61 NNCSSITSGYPCCSNNCKVEYTDN----DGKWGVENNWNWCGISNCGGGQQOQPTOP 116
QY 240 VSGLARFPQPTPYDLDLLDFQQALEQNGHYSGASVQADFRLQGD-----RVPVK 290
Db 117 TQPTQPOOPTOPSSDNF---FENEIYSNYKFGQ-EVDISIKKLGDLKAKAEKYVYPTA 172
QY 291 V-----SVTEVRKHLETKIR-----LDSEYGLGKIAVDYNNLENKGYIG 331
Db 173 VWLAWDGAPOEVPYRLQEAQNKTVFVLYMPTRDCCANASAGGSATIDKY----KGYIN 228
QY 332 SVVWMDKYETTLAAGISQPRNYRGNYWTSTNLEKRAFSGGIWYVRDRAGI 391
Db 229 NIYNTSNQYKNSKIVMILEPDTI--GNLVNTN--NDCNRVNRNMHKOALSYSKFGTQSHV 286
QY 392 DARLGAEFLAEGRKIPGSDIDLGNSHATMLTASWKROLLNNVLPENGHYLDGKIGTTLG 451
Db 287 KVLDA---AHGAWL-----NOYADQ-TANVIKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKGTLFIIRG-QAGYTVARDNADVPGLMFRSGGASS 510
Db 326 TNVSNYOSIESEYK---YHQNLRALSKGVRGLKFIVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLQORPKGNPNPGSMP 401

RESULT 37
US-08-613-009A-10
; Sequence 10, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
```

```
RESULT 36
US-09-687-147-2
; Sequence 2, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-687-147-2

Query Match 3.4%; Score 108; DB 4; Length 459;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 90; Conservative 63; Mismatches 193; Indels 94; Gaps 19;

QY 141 ILGDILSDGNLAEYRNALENWQOPVGS-----FDQDSWNSK 179
Db 6 VLATLFATGALASEC-----HWQYPCCKDCTVYVYTDTEGKWGLNNDWCMIDNRCSSNN 60
QY 180 TSVLGAVTRKGYPLAKIGNTRAAYNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQI 239
Db 61 NNCSSITSGYPCCSNNCKVEYTDN----DGKWGVENNWNWCGISNCGGGQQOQPTOP 116
QY 240 VSGLARFPQPTPYDLDLLDFQQALEQNGHYSGASVQADFRLQGD-----RVPVK 290
Db 117 TQPTQPOOPTOPSSDNF---FENEIYSNYKFGQ-EVDISIKKLGDLKAKAEKYVYPTA 172
QY 291 V-----SVTEVRKHLETKIR-----LDSEYGLGKIAVDYNNLENKGYIG 331
Db 173 VWLAWDGAPOEVPYRLQEAQNKTVFVLYMPTRDCCANASAGGSATIDKY----KGYIN 228
QY 332 SVVWMDKYETTLAAGISQPRNYRGNYWTSTNLEKRAFSGGIWYVRDRAGI 391
Db 229 NIYNTSNQYKNSKIVMILEPDTI--GNLVNTN--NDCNRVNRNMHKOALSYSKFGTQSHV 286
QY 392 DARLGAEFLAEGRKIPGSDIDLGNSHATMLTASWKROLLNNVLPENGHYLDGKIGTTLG 451
Db 287 KVLDA---AHGAWL-----NOYADQ-TANVIKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKGTLFIIRG-QAGYTVARDNADVPGLMFRSGGASS 510
Db 326 TNVSNYOSIESEYK---YHQNLRALSKGVRGLKFIVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLQORPKGNPNPGSMP 401

RESULT 37
US-08-613-009A-10
; Sequence 10, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
```



```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:1b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 702 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-867-941-25

Query Match      3.4%; Score 107.5; DB 2; Length 702;
Best Local Similarity 20.2%; Pred. No. 0.42;
Matches 112; Conservative 69; Mismatches 202; Indels 171; Gaps 26;

QY 33 GFALFKNKSPDTSVKLKPKFPVRIDTQDSEIKDMVEEHLPLITQQQEEVLDKEQTGFLA 92
Db 33 GFALFKNKSPDTSVKLKPKFPVRIDTQDSEIKDMVEEHLPLITQQQEEVLDKEQTGFLA 92
QY 103 GMALSKINLHNRDTPDDEKNIITLQK-KQVAGKSKPLPFLSDVFNKLLD-----GYIA 157
Db 103 GMALSKINLHNRDTPDDEKNIITLQK-KQVAGKSKPLPFLSDVFNKLLD-----GYIA 157
QY 93 EAPDNVKTMLRSKGYFSSKVSUTEKDAYTVHTPQPTKIANVGVAIIGDILSDGNLA 152
Db 93 EAPDNVKTMLRSKGYFSSKVSUTEKDAYTVHTPQPTKIANVGVAIIGDILSDGNLA 152
QY 158 -----KMNVDKNAI-----GDRIKKGKKEI-----SDEELA 184
Db 158 -----KMNVDKNAI-----GDRIKKGKKEI-----SDEELA 184
QY 153 EYRNALENHQOPVGSDFDQ--DSWEN-----SKTSVLGAVTRK-----CYPLAKLGNT 200
Db 153 EYRNALENHQOPVGSDFDQ--DSWEN-----SKTSVLGAVTRK-----CYPLAKLGNT 200
QY 185 KQKEAVRK-----SHEFOQVLSLENKIPHSNDGTTKATTRDLKYVDYGYLLANDGN-- 237
Db 185 KQKEAVRK-----SHEFOQVLSLENKIPHSNDGTTKATTRDLKYVDYGYLLANDGN-- 237
QY 201 AAVNPOTATVDLNVVDSGRPIAFGDEITGTQRYPEQIVSGLARFQPGTPYDLDLDD 260
Db 201 AAVNPOTATVDLNVVDSGRPIAFGDEITGTQRYPEQIVSGLARFQPGTPYDLDLDD 260
QY 238 -----YLTVKTDKLNGLG-PV--GGVYNGTTAKE-----LPT 268
Db 238 -----YLTVKTDKLNGLG-PV--GGVYNGTTAKE-----LPT 268
QY 261 QQALEONGHYSGASVQADFRLQGDVVPKVSQVTEV-KRKLETGIRLDSEYGLGCKIAY 319
Db 261 QQALEONGHYSGASVQADFRLQGDVVPKVSQVTEV-KRKLETGIRLDSEYGLGCKIAY 319
QY 269 QDAVKYKGHW-----DF-----MTDVANRRNRSEVKENSQAGW----- 302
Db 269 QDAVKYKGHW-----DF-----MTDVANRRNRSEVKENSQAGW----- 302
QY 320 DYINLKNKGYIGSVVWMDKYEITLAAGISOPRNYRCNYWTSNVSYNRSTTONLEKRAFS 379
Db 320 DYINLKNKGYIGSVVWMDKYEITLAAGISOPRNYRCNYWTSNVSYNRSTTONLEKRAFS 379
QY 303 -YYGASSK-----DEYNRLTKEDSAPDGHSGEY-----GHSSEFTVNFKEKLLT 346
Db 303 -YYGASSK-----DEYNRLTKEDSAPDGHSGEY-----GHSSEFTVNFKEKLLT 346
QY 380 GGIWYVRDRAGIDARLGAEPFAEGRKIPGSDIDL---GNSHATMLTASWKROLNNVLP 436
Db 380 GGIWYVRDRAGIDARLGAEPFAEGRKIPGSDIDL---GNSHATMLTASWKROLNNVLP 436
QY 347 GKLF-----SNLQDRHKGNVTKTERYDIDANTHGNFRGCSATASKNKNTSKHPFTS 397
Db 347 GKLF-----SNLQDRHKGNVTKTERYDIDANTHGNFRGCSATASKNKNTSKHPFTS 397
QY 437 ENGHYLDGKGTGTLTGLSTALIRTSARAGYEFTPENKKLGTFTIIRGQAGYTVARDNA- 495
Db 437 ENGHYLDGKGTGTLTGLSTALIRTSARAGYEFTPENKKLGTFTIIRGQAGYTVARDNA- 495
QY 398 DANNRLEGGFYGPKEBEL-----AGKELTNDNKLFGVF---GAKRESKAEEKTE 443
Db 398 DANNRLEGGFYGPKEBEL-----AGKELTNDNKLFGVF---GAKRESKAEEKTE 443
QY 496 ---DVPSGLMFRSGGASSVRGY---ELDSIGLAGPNSVLPERALLVGSLEYOLPFTRTL 549
Db 496 ---DVPSGLMFRSGGASSVRGY---ELDSIGLAGPNSVLPERALLVGSLEYOLPFTRTL 549
QY 444 AILDYALGTFNTSNATFTTPEFTEKQLDNFGNA-----KKVLGSTVIDLVPTDAT 494
Db 444 AILDYALGTFNTSNATFTTPEFTEKQLDNFGNA-----KKVLGSTVIDLVPTDAT 494
QY 550 SGAVFHDMDGDAAN 563
Db 550 SGAVFHDMDGDAAN 563
QY 495 KNEFTKDKPESATN 508
Db 495 KNEFTKDKPESATN 508
```

Search completed: November 9, 2002, 01:21:40
Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:17:09 ; Search time 21 Seconds
(without alignments)
2787.898 Million cell updates/sec

Title: US-09-857-669-2
 Perfect score: 3173
 Sequence: 1 MWIKPTALLPALFFPHAY.....IAYGSDKKIRNHIISLGTRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 2832224

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database :      PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	3153	99.4	615	2	A82025	probable outer mem	
2	3136	98.8	535	2	G81003	conserved hypothe	
3	449	14.2	579	2	F83327	conserved hypothe	
4	379.5	12.0	577	2	S56445	hypothetical 64.8K	
5	379.5	12.0	577	2	F91278	hypothetical prote	
6	379.5	12.0	577	2	F86119	hypothetical prote	
7	374	11.8	578	2	B64012	hypothetical prote	
8	368.5	11.6	578	2	A60428	probable exported	
9	355.5	11.2	577	2	A10555	probable exported	
10	348	11.0	582	2	F82064	conserved hypothe	
11	323.5	10.2	641	2	F97672	conserved hypothe	
12	320	10.1	617	2	AF2897	conserved hypothe	
13	306.5	9.7	617	2	G82706	conserved hypothe	
14	290.5	9.2	623	2	A13488	conserved hypothe	
15	279	8.8	833	2	AF2089	outer membrane pro	
16	272.5	8.6	628	2	B87448	hypothetical prote	
17	253	8.0	797	2	D82000	conserved hypothe	
18	253	8.0	797	2	G81228	outer membrane pro	
19	246	7.8	778	2	C70412	outer membrane pro	
20	246	7.8	795	2	AC0129	probable surface a	
21	227	7.2	803	2	B82099	surface antigen VC	
22	226	7.1	803	2	AB0530	outer membrane pro	
23	217	6.8	810	2	A64742	outer membrane pro	
24	217	6.8	810	2	C90651	hypothetical prote	
25	217	6.8	810	2	C85502	hypothetical prote	
26	215.5	6.8	797	2	H83190	hypothetical prote	
27	211	6.6	576	2	A24117	probable outer mem	
28	210.5	6.6	769	2	F87486	hypothetical prote	
29	209.5	6.6	797	2	JC4078	outer membrane pro	
						protective surface	

ALIGNMENTS

RESULT 1

RESUL
A82025

AB020253
probable outer membrane protein NMA02996 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB020253
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M
R; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
Nature 404, 502-506, 2000

[illegible]

QY	241	SGLAREQPCTPYDLDLLDFQOALEQNGHYSGASVQADFDRLOGQDRVPVKVSVTEVKRH	300
Db	247	SGLAREQPCTPYDLDLLDFQOALEQNGHYSGASVQADFDRLOGQDRVPVKVSVTEVKRH	306
QY	301	LETGIRLDSYGLGGKIADYYNLFNKKYIGSVVWDMDKYETTLAAGISQPRNYRGNYWT	360
Db	307	LETGIRLDSYGLGGKIADYYNLFNKKYIGSVVWDMDKYETTLAAGISQPRNYRGNYWT	366
QY	361	SNVSYNRSYTNQNLKRAFSGGIWYVDRDAGIDARLGAELAGRKIPGSDIDLGNSHATM	420
Db	367	SNVSYNRSYTNQNLKRAFSGGIWYVDRDAGIDARLGAELAGRKIPGSDIDLGNSHATM	426

Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSTALIRTSARAGYFFTPENKKLGT 480
Db 427 LTASWKROLLNNVLPENGHYLDGKIGTTLGAFSTLIRTSARAGYFFTPENKKLGT 486
Qy 481 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNCVSLPERALLVGSLE 540
Db 487 IIRGAGYTVARDNANVPGLMFRSGGASSVRGYELDSIGLAGPNCVSLPERALLVGSLE 546
Qy 541 YOLPPTRTLGSAGVAFHDMGDAANFKRMKLKHGSGLGVRWFSPFLAPFSFDIAYGHSDDKKIR 600
Db 547 YOLPPTRTLGSAGVAFHDMGDAANFKRMKLKHGSGLGVRWFSPFLAPFSFDIAYGHSDDKKIR 606
Qy 601 WHISLGRF 609
Db 607 WHISLGRF 615

RESULT 2
C81003
conserved hypothetical protein NMB2134 [Imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81003
R:Tetelin, H.; Saunders, N.J.; Heideberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Vitti, A.; et al.
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: G81003
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-635 <TET>
A:Cross-references: GB:AB002561; GB:AB002098; NID:g7227384; PIDN:AAF42442.1; PID:g722739
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB2134

Query Match 98.8%; Score 3136; DB 2; Length 635;
Best Local Similarity 98.7%; Pred. No. 1.2e-202;
Matches 601; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MWIKPTALLPFFHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIOTQ 60
Db 27 MWIKPTALLPFFHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVLIOTQ 86
Qy 61 DSEIKDMVEEHLPLITQOEEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLETKDG 120
Db 87 DSEIKDMVEEHLPLITQOEEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLETKDG 146
Qy 121 AYVHTTPGRTKIANGVAILGDLSDGNLAEEYRNALNNQWQPVGSDFDQDSWENSKT 180
Db 147 AYVHTTPGRTKIANGVAILGDLSDGNLAEEYRNALNNQWQPVGSDFDQDSWENSKT 206
Qy 181 SVLGAVTRKGYPLAKGNTRAAVNPDTATVLDNVVDSGRPIAGDFEITGTQRYPEQIV 240
Db 207 SVLGAVTRKGYPLAKGNTRAAVNPDTATVLDNVVDSGRPIAGDFEITGTQRYPEQIV 266
Qy 241 SGLARFQPGTVDLLDLFQALQNGHYSGASVQADFRLQDGRVPVKVSVTEVKRKH 300
Db 267 SGLARFQPGTVDLLDLFQALQNGHYSGASVQADFRLQDGRVPVKVSVTEVKRKH 326
Qy 301 LETGIRLDSYGLGCKIAYDYNLFNKGYSVVMVMDKYEETTLAAGISQPNRYGNWT 360
Db 327 LETGIRLDSYGLGCKIAYDYNLFNKGYSVVMVMDKYEETTLAAGISQPNRYGNWT 386
Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLAEGRKIPGSDIDLGNSHATM 420
Db 387 SNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLAEGRKIPGSDIDLGNSHATM 446
Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSTALIRTSARAGYFFTPENKKLGT 480

Db 447 LTASWKROLLNNVLPENGHYLDGKIGTTLGTFLSTALIRTSARAGYFFTPENKKLGT 506
Qy 481 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNCVSLPERALLVGSLE 540
Db 507 IIRGAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNCVSLPERALLVGSLE 566
Qy 541 YOLPPTRTLGSAGVAFHDMGDAANFKRMKLKHGSGLGVRWFSPFLAPFSFDIAYGHSDDKKIR 600
Db 567 YOLPPTRTLGSAGVAFHDMGDAANFKRMKLKHGSGLGVRWFSPFLAPFSFDIAYGHSDDKKIR 626
Qy 601 WHISLGRF 609
Db 627 WHISLGRF 635

RESULT 3
F83327
conserved hypothetical protein PA2543 [Imported] - *Pseudomonas aeruginosa* (strain PAO)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83327
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.; Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83327
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AB004682; GB:AB004091; NID:g9948598; PIDN:AAG05931.1; GSPDB:G
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA2543

Query Match 14.2%; Score 449; DB 2; Length 579;
Best Local Similarity 22.9%; Pred. No. 2.7e-22;
Matches 144; Conservative 111; Mismatches 289; Indels 84; Gaps 16;

Qy 9 LLPALFRPHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIOTQDSEIKDMV 68
Db 3 LTQGLFRF-----PLAGLCLSTAAFA-----ESQLSVRTTANAALKANI 43
Qy 69 BEHLPLITQOEEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLETKDG---AYTVH 125
Db 44 EAYVSGSLGERDEAALQ-----FRNNAEQAQAALGYFQAQIDSEVKDGKPKLTK 98
Qy 126 ITPGRTKIANGVAILGDLSDGNLAEEYRNALNNQWQ-----VGSDFDQDSWENSK 179
Db 99 VVPGEVRLQVNIQVLGEAAS-----LESFRLPSGKQLKPGAKLNQGVYEDAK 147
Qy 180 TSVLGAVTRKGYPLAKGNTRAAVNPDTATVLDNVVDSGRPIAGDFEITGTQRYPEQI 239
Db 148 RLIONQASRYGTFQGRFSTQRLSDIPRAGIADIDLVDGQGYTFGKVSFGDSDIEEL 207
Qy 240 VSGLARFQPGTVDLLDLFQALQNGHYSGASVQADFRLQD-----RVPVKVSVTE 295
Db 208 LRRVVPFRKAGPYDELIAELNQLNQLSSGYFEGVRVDAAPQAQADGARQAIIPAVRLEA 267
Qy 296 VKRIKLETGIRLDSYGLGCKIAYDYNLFNKGYSVVMVMDKYEETTLAAGISQPNRYGNWT 341
Db 268 RKPRTMGVGLGTFDVGARARFNTRHWNNAEGHSLGSEISAPRONVGA--W-----YE 321
Qy 342 TTAAGISQPNRYGNWTSMVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLA 401
Db 322 IPLDPLTDLKURFSGYQFEDLVDTESKLLTL-----GGEMHSKRDPGWORVSVSLNMR 375
Qy 402 EGRKIPGSDIDLGNSHATMLTASWKROLLNNVLPENGHYLDGKI-GTTLGTFLSTAL- 459
Db 376 EYKLGSD--DSGLSSFLMPGIGYSLLTENDKDPHGYRLOFNVKGAKEGLLADADVLH 432
Qy 460 IRTSARAGYFFTPENKKLGTETIRGQAGYTVARONADVPGLMFRSGGASSVRGYELDSI 519

Db 433 VDAKAGLTTFAGCHRLG-----RLQVGGIATNDYKIPSLRPFAGGDOSVRGYDRTL 489
QY 520 GLAGPNSVLPERRALLVGSLEYQLPFTRTLSGAVFHDMDGAAAFKMKLKHGSLGVWR 579
Db 489 SPENSDDKIGGRYMIAGSVEYQYPLAERMLATFVDDGNAFNSLDSPSIKTGVGFCVR 548
QY 580 FSLAPFSFDIAG-HSDKKIRWHISLG 606
Db 549 VSPVGPLRLDLAHLADDGGRFLHFSMG 576

RESULT 4

S56445
hypothetical 64.8k protein (msra-chpbi intergenic region) - Escherichia coli (strain K-12)
N:Alternate names: hypothetical protein c577
C:Species: Escherichia coli
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C:Accession: S56445; G65233
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Accession: S56445
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-577 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97116.1; PID:g537061

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G65233
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-577 <BLAT>
A:Cross-references: GB:AE000493; GB:U00096; NID:g2367360; PIDN:AAC77177.1; PID:g1790666;
A:Experimental source: strain K-12, substrain MGL655

C:Genetics:
A:Gene: ytfM
A:Start codon: GTG

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

QY 74 LITQOQEVLDKEQTGLAEAPDNVKTMLRSKGYFSSKVS-----TEKDGYTV--HIT 127
Db 43 LSTIESDEVTPDDR---FRARVDDAIRGLKALGYOPTYEFDLRPPPKGRQVLIKAVT 99

QY 128 PGPRTKIANYGVAIIGDILSDGNLAERYRNALENWQOPVGSDFDQDSDWENSKTSVLGAVT 187
Db 100 PGVPVLIGTDVVLRGARTD---KDYLK--LLDTRPAIGTVLNOGYENFKSLSIAL 154

QY 188 RKG-----PLAKIGNTRAADVNPDTATVDLNVVVDGSRPIAFGDFEITGTQRYPEQIVSG 242
Db 155 RKGYDFSEFTKAQLG---IALGLHKAFWID--YNSGERYRFGHVTEGSGQ-IRDEVLYQN 208

QY 243 LARPOPTPTDLDLLDFOQALEONGHYSASVQADFDRLQGDV--PKVSVTEVKRHL 301
Db 209 LVPPKEGDEYKDLAELNRLSATGWNFSVVAPOFDKARETKVLPVTGVSPTENT 268

QY 302 ETGIRLDSYGLGKIAIDYNNLFNKGYIGSVVWMDKYEITTLAAGISOPRNRGNWYTS 361
Db 269 ETGVGYSTDVGPVKATW-----KKPMNSY-----CHSLTSTISAPQ-----TL 311

QY 362 NVSYNRSTTONLEKRAFSGGIWYVDRAGIDARLGAELFAEGRKIPGSDIDLGNHATML 421
Db 312 DFSYKMPLLKN-----PLEQYLYVQG--GFKRTLDNDTSDSTTL 349

QY 422 TAS-----WKQLNNVLPENGHYLDGKIGTTLTGLSTLSTALRTSARAGYF----- 469

Db 350 VASRYWDLSSGWORAI---NLRWSLDHTQGEITNTTLMFLFPGVMISRTSRGGLMPTWG 406
QY 470 -----FTPENKKLGT-----FIIRGOAGYTVARNADVPDGL 501
Db 407 DSQRYSDYDNTANGSDVDFSVFOAQNWIIITLYDRHRFVTRGTLGWIETGDFDKVPPDL 466
QY 502 MFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSGAVFHDMDGAA 561
Db 467 RFFAGGDRSIRGYKYKSIAPKYANGDLKGAKLITGSLEYQYNYVTGKWWGAVFVDSGEAV 526
QY 562 ANFKRMKLKHGSLGVRFHFSPLAPFSFDIAGYHSDKK---IRWHISLG 606
Db 527 SDIRSRDFKTGTGVGRWESVPGVPIKLDFAVPVADKDEHGLQFYIGLG 574

RESULT 5

F91278

hypothetical protein Ec5198 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F91278

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: F91278
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-577 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA38621.1; PID:g13364675; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:
A:Gene: Ec5198

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

QY 74 LITQOQEVLDKEQTGLAEAPDNVKTMLRSKGYFSSKVS-----TEKDGYTV--HIT 127
Db 43 LSTIESDEVTPDDR---FRARVDDAIRGLKALGYOPTYEFDLRPPPKGRQVLIKAVT 99

QY 128 PGPRTKIANYGVAIIGDILSDGNLAERYRNALENWQOPVGSDFDQDSDWENSKTSVLGAVT 187
Db 100 PGVPVLIGTDVVLRGARTD---KDYLK--LLDTRPAIGTVLNOGYENFKSLSIAL 154

QY 188 RKG-----PLAKIGNTRAADVNPDTATVDLNVVVDGSRPIAFGDFEITGTQRYPEQIVSG 242
Db 155 RKGYDFSEFTKAQLG---IALGLHKAFWID--YNSGERYRFGHVTEGSGQ-IRDEVLYQN 208

QY 243 LARPOPTPTDLDLLDFOQALEONGHYSASVQADFDRLQGDV--PKVSVTEVKRHL 301
Db 209 LVPPKEGDEYKDLAELNRLSATGWNFSVVAPOFDKARETKVLPVTGVSPTENT 268

QY 302 ETGIRLDSYGLGKIAIDYNNLFNKGYIGSVVWMDKYEITTLAAGISOPRNRGNWYTS 361
Db 269 ETGVGYSTDVGPVKATW-----KKPMNSY-----CHSLTSTISAPQ-----TL 311

QY 362 NVSYNRSTTONLEKRAFSGGIWYVDRAGIDARLGAELFAEGRKIPGSDIDLGNHATML 421
Db 312 DFSYKMPLLKN-----PLEQYLYVQG--GFKRTLDNDTSDSTTL 349

QY 422 TAS-----WKQLNNVLPENGHYLDGKIGTTLTGLSTLSTALRTSARAGYF----- 469

Db 350 VASRYWDLSSGWORAI---NLRWSLDHTQGEITNTTLMFLFPGVMISRTSRGGLMPTWG 406

QY 470 -----FTPENKKLGT-----FIIRGOAGYTVARNADVPDGL 501

Db 407 DSQRYSDYDNTANGSDVDFSVFOAQNWIIITLYDRHRFVTRGTLGWIETGDFDKVPPDL 466

QY 502 MFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSGAVFHDMDGAA 561

Db 467 RFFAGGDSIRGYKYKSTAPKYANGDLKGASKLINGSLEYQYNVTKWVGAVFVDSGEAV 526
Qy 562 ANFKRMKLKHGSGLVRFSPFSLAPSFDIAYGHSDDK---IRWHISLIG 606
Db 527 SDIRSDFKTGTGCVGRWESPVGPKLDFAVPVADKDEHGLQFYICLG 574

RESULT 6
F86119
hypothetical protein ytfm [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F86119
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, M.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F86119
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <STO>
A:Cross-references: GB:AE005174; NID:gl2519219; PIDN:AAG59418.1; GSPDB:GN00145; UWGP:Z58
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ytfm

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

Qy 74 LITOOQEEVLDEQGTGLAEAPDNVKTMLRSKGYFSSKVSLSL---TEKDGAYTV--HIT 127
Db 43 LSTIESDEVTPDR--FRARVDAAIREGLKALGYQPTIEFDLPPPKGRQVLIKVT 99

Qy 128 PGPRTKIANVGVAIGDILSDGNLAEYRNALNNQVPQVDSFDDSDSWENSKTSVLGAVT 187
Db 100 PGVPVLIGGTDVLRGGARTD---KDYLK--LLDTPAIGTVLNGQDYNFNKKSITSL 154

Qy 188 RKGYS----PLAKGNTRAAVNPDTATVLDNVVDSGRPIAFGDFEITGTORYPEQIYSG 242
Db 155 RKGYPDSFETRAQLG---IALGLHKAFWID--YNSGERYFGHVTPEGSQ--IRDEYLN 208

Qy 243 LARFOPGTPYDLDDLDDLLDQOALEONGHYSGASVQADFRLQGDYV-PYKVSYTEYKRHL 301
Db 209 LVPPFKEGDEYESKDLAELNRLSATGWFNSVVAPOFDKARETKVLPLTGVVSPRTNTI 268

Qy 302 ETGIRLDSYGLGGKIADYDYNLFNKGYIGSVVWMDKRYETTLAAGISQPRNRYGNWTS 361
Db 269 ETGVGYSTDVGPVRKATN-----KKPMNSY-----GHSLTSTSIAPEQ-----TL 311

Qy 362 NVSYNRSTQNLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATML 421
Db 312 DSYKMWPLLN-----PLQYLYVQG-GFKRTDLNDESDSTL 349

Qy 422 TAS-----WKROLLNNVLPENGHYLDCKIGTTLGTFLSSTALIRTSARAGYF---- 469
Db 350 VASRYWDLSSGWRRAI---NLRWSLDHFTQGEITNTMLFYPGVMSRSTRSGGLMPTWG 406

Qy 470 -----FTPENKKLGT-----FIRGQAGYTVARNADVPSSL 501
Db 407 DSQRYSDYSNTAMGSDVDFSVFQAQNVWIRFLYDRHFRVTRGTGLGTETGDFDKVPPDL 466

Qy 502 MFRSGGASVRYGLSDISGLAGPNSGVLPERALLVGSLEYQLPFTRLTSGAVFHDMDGAA 561
Db 467 RFFAGGDSIRGYKYKSTAPKYANGDLKGASKLINGSLEYQYNVTKWVGAVFVDSGEAV 526

Qy 562 ANFKRMKLKHGSGLVRFSPFSLAPSFDIAYGHSDDK---IRWHISLIG 606
Db 527 SDIRSDFKTGTGCVGRWESPVGPKLDFAVPVADKDEHGLQFYICLG 574

RESULT 8

RESULT 7
B64012
hypothetical protein HI0698 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64012
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vento
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64012
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-578 <TIGR>
A:Cross-references: GB:U32752; GB:U42023; NID:gl1573692; PIDN:AAC22357.1; PID:gl1573700

Query Match 11.8%; Score 374; DB 2; Length 578;
Best Local Similarity 22.9%; Pred. No. 2.9e-17;
Matches 153; Conservative 93; Mismatches 263; Indels 158; Gaps 18;

Qy 3 IKPTALLPALPFFPHAYAPADLSENKAAGFALEFKNKSPTDESVKLAKPKPVRIDTODS 62
Db 6 LKLTALFL-ALSCPPAFABEQTVDIEVOGIRFRAVRMTD-----LNVNLINK 51

Qy 63 ETKOMVEEHLPLITOOQEEVLDEKQGTGLAEAPDNVKTMLRSKGYFSSKVSLSLEKDO--- 119
Db 52 EEMDGSERYQHLVTKA-----VDRGLRVFGYESSVRFERKQRQG 91

Qy 120 --GAYTVHITGPRTKIANVGVAIGDILSDGNLAEYRNALNNQVPQVDSFDDSDSWEN 177
Db 92 KRDLIAHVTGPCEPTKIAGTDVQIEGAAQDENF-----NALRKNLKPXGVLVEHQTYDD 146

Qy 178 SKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVLDNVVDSGRPIAFGDFEITGTORYPE 237
Db 147 YKTAISRLALNARGYPDGNFKISRLIEISPETHQAMRMFLFDGSRVYHYGNITFHSQ--IRD 205

Qy 238 QIVSGLARFPQCTPYDLDLLDFOQALEONGHYSCASVQADFRLQGDYVYPKVSYTEVK 297
Db 206 DYLNILNLIKSGDPYLMNLSLDTSDFFSSNWFSSVLVQPNVNH-KSKTVDVEIILYPRK 264

Qy 298 RHKLETGIRLSEYGLGKIAIY-----319
Db 265 KNAMELGVGFTDGGVHGIGWTGKPINSRGHSLSRLNLYLSAPKQTLKATYRMLLKNPL 324

Qy 320 DYNNLFNKGYIGSVVWMDKRYETTLAAGISQPRNRYGNWTS-----NVSYNRST 369
Db 325 NYIYDFAVGWEKEKENDTNTRVLITLSA-----LRYMNAHGWQYFGGLRMRYDSFT 375

Qy 370 TONLEKRAF---SGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASW 425
Db 376 QADITDKTLLLYPTVGFTRTLRGGSFATWG-----DVQKI---TFDLS----- 416

Qy 426 KRQLNNVLPENGHYLDCKIGTTLGTFLSSTALIRTSARAGYFT-PENKKLGTPIRG 484
Db 417 KR-----IWLSESSFVKVQASSAWVRTYAENHRV---VARA 449

Qy 485 QAGYTVARNADVPSSLMFRSGGASVRYGLSDISGLAGPNSGVLPERALLVGSLEYQLP 544
Db 450 EIGYLHTGKIEKIPPTLRFPPAGGDSRVRGYGYKKLAPKNRNKGLVGGSKRLTTSLEYQIQ 509

Qy 545 FTRTLGSAVFHDMDGAAANFKRMKLKHGSGLVRFSPFSLAPSFDIAYGHSDDK---KKIR 600
Db 510 VYPNMAATFADSGLAADNYATAKELRYGTGCVGRWASPVGAIKFDIATPIRDKNSKNIQ 569

Qy 601 WHISLGT 607
Db 570 FYIIGLT 576

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82064
R:Hidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P. L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <HEI>
A:Cross-references: GB:AE004323; GB:AE003852; NID:g9657119; PIDN:AAF95689.1; GSPDB:GN00101
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2548
A:Map position: 1

Query Match 11.0%; Score 348; DB 2; Length 582;
Best Local Similarity 23.1%; Pred. No. 1.6e-15;
Matches 136; Conservative 96; Mismatches 255; Indels 102; Gaps 19;

Qy 65 KDMVEHLPLITQOO-----EVLDKQGTFLAEAPDNVKTMLRSKGYFSSKVSUT- 116
Db 46 KDNVEAYLSSIAAQDYSTSLRFQSLERSMT-----EA-----LNALGYHPSIDFTV 93

Qy 117 -EKDGYTVHTTPGRTKIANVGVAIGDLSDGNLAEEYRNALNQOPVGSDFDQDSW 175
Db 94 SEDNORLRAAVLGEVTRLSEVOIVIRGEAGDRDORLRRS-----GLRVADPLNLSY 149

Qy 176 ENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORY 235
Db 150 DNLKSGTRNLALQGYNGDFQASRLEVIPELNQARVILHFDGIRYLFGATTVEGSO-I 208

Qy 236 PEQIVSGLARPOGPTDYDLDDLLDFOQALBONGHYSGASVOADFDRL-QGDRVPVKVSVT 294
Db 209 DENRVMSLRPKQGEPLVSOVGFNQNLSTNDFSSVFVEPDLSQLDEGRELPKVTLA 268

Qy 295 EVKHKLETGIRLSEYGLGKIAIDYNNLFNKYIGSVVMDKDYETTLAAGISQP--- 351
Db 269 POARNQLETGLGYSTDGVGRSLKWKPKWNSQGHSDFSFSLSEIPQITAGYKIFLED 328

Qy 352 ---RNYRGNVTSNVSNNRSTONLEKRAFSGGIWYVDRAGIDARLGAEEFLAEGRKIPG 408
Db 329 ALNEYRYIQYGMKHL--DKRDTESLNSLERHW----- 361

Qy 409 SDIDLGNHATMLTASWKRLNNV---LHPENGY-LDGKIGTTLGT----- 452
Db 362 -QLD-GGWHRTVFT----RYLLENYRQGLQDQNSQFLPLPGMTYRTRTRNSGILLTWGDK 415

Qy 453 -----FLSSTALIRTSARAGYFTT-PENKILGTFTIRGOAGYTVARDNADVPSG 500
Db 416 QTITLEYGDALLSETVRLQTOGSSWLRTVARNHRA---LVRVDGANLVDEFDQLSPS 472

Qy 501 LMFRRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRLTSGAVPHMDGA 560
Db 473 LRFAGGDNRLRGYKYSISPDQASGALTGAKYIATSSIEYQYRLTGNWAAEMFMDVGA 532

Qy 561 AANFKRMLKHGSLGVRWFSPLAPFSDIAYG---HSDKIRWHISLG 606
Db 533 FND--NPENKMGVGTGRWISVPGPIRLDFAWGLDAAPGDEFKIHFTIG 579

RESULT 11
H97672
hypothetical protein AGR_C4742 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97672
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markeiz, B.;
Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H97672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-641 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88337.1; PID:g15157817; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C4742
A:Map position: circular chromosome

Query Match 10.2%; Score 323.5; DB 2; Length 641;
Best Local Similarity 22.6%; Pred. No. 8.4e-14;
Matches 148; Conservative 94; Mismatches 290; Indels 123; Gaps 21;

Qy 6 TALLLPALFFPHAYAPAAADLSENKAAGFALFRNKSDDTESVKLPKFPVRIDTQDSE-- 63
Db 19 TALAVALTFALYPAP--ARDAFAFKLFGRMLGSGEEVEVIN-PVKYAVTLOAADAKS 75

Qy 64 IKDMVEHLPLITQOOEVLDKQGTFL-----AEEAPDNVKTMLRSKGYFSSKVSUT-- 116
Db 76 LKSSLENSLLADK-----DKPASGDLGLLIKARDDRDLIAALYENARYGGIVNVTV 130

Qy 117 -----EKDGYTVHTTPGRTKIANVGVAIGDLSDGNLAEEYRNALN 161
Db 131 GKNVDDLPPNPVFDHSTPVMVITVTPGPKFTLGN--VRLEGDV-TCRNLDEYGLIA--- 184

Qy 162 WQPVGSDFDQDQSHENSKTSVLGNVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRP 221
Db 185 -----GGDAGSLAIRAGNKLIDDLKAERPLAKLTREAVANHATNTVDITMAAEGVP 239

Qy 222 IAFGDPEITGTQRYPEQIVSGLARFPQPTPYDILLDDLQOALBONGHYSGASVOADFDR 281
Db 240 APIGAVTVTEKTVGDFIRYRSLNGCEPSPKLRKAADRLRLQLGVFSSLTIKEAGTL 299

Qy 282 LQDRVPVKVSVTEVKRHKLETGIRLSEYGLGKIAIDYNNLFNKG-----YIGSV---- 333
Db 300 ARDGTIPLTIEVSGKHYRFGVGAQYSTTEGIGLQGYGHRNLFPGAESLRIGESVSRIA 359

Qy 334 ---VMDMD-----KYETTLAGISQPRNYRGNVTSNVSNNRSTQN 372
Db 360 EASVEGMDYSAGITFTKPGMNFNRTFTKTSLIAKTENPDYTRAKTLTGT----- 409

Qy 373 LEKRAFSGGIWYVDRAGIDARLGAEEFLAEGRKIPGSDIDLGNHATMLTASWKRLNN 432
Db 410 -----AGFAVELND-----TDTAAGLEVQWADTEDAFCKNEYLTITSIPLEFVRD 454

Qy 433 VLHPENGYLDGKIGTTLGTFLSSTALIRTSARAGYFTTPE-----KKLGT---PIIR 483
Db 455 T-----RDKLNPTEGFRASLAAPKPSVEALNGTFESSPEGSITGYKGLGAEDRLIMA 506

Qy 484 GQ--AGYTV-ARONADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLE 540
Db 507 GLKSGGVLVGSDQLDIPITTRRFAGGGGSGVRGYOIEISPYNAAGDAGTGRSVYVGSVE 566

Qy 541 YQLPFTRLTSGAVPHDMG---DAANFKRMLKHGSLGVRWFSPLAPFSDIA 591
Db 567 ARIKVTTDITGLVPFEDAGVYSDGYTPDPFS--DIRAGAGIGLRYATPGPLRLDVA 619

RESULT 12
AF2897
conserved hypothetical protein Atu2615 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2897
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, C.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Bliddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AF2897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-617 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL43596.1; PID:gl7741114; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu2615

A:Map position: circular chromosome

Query Match 10.1%; Score 320; DB 2; Length 617;
Best Local Similarity 22.4%; Pred. No. 1.4e-13;
Matches 147; Conservative 92; Mismatches 286; Indels 130; Gaps 21;

```
QY 6 TALLPALFFPHAYAPADLSENKAAGPALFKPKNSPDOTESVKLPKFPVRIDTODSE-- 63
DB 2 TTYLPAPF-----ARDAPFLFGMLWGSEPEVEVIN-PVKYAVTLDAAADKS 51
QY 64 IKDMVEHLPLITQOOEEVLDKEQTGL-----AEEAPDNVKTMLRSKGYFSKVSILT-- 116
DB 52 LKSSLENSLLADK-----DRPASGDLGLLIKARDDRDLTAALYENARYGGIVNVYA 106
QY 117 -----EKDGAIVVHTTPGPRTKIANVGVAIGDILSDGNLAEEYRNALEN 161
DB 107 GKNVDDLPPNPVFDHSTPVPMTVTPGPKFTLGN--VRLEGDV-TGRNLDEYGLIA--- 160
QY 162 WQOPVGSDFDQDSWENSKSVLCAVTRKGYPLAKLGNTRAAVNPDPATVDLNVVDSGRP 221
DB 161 -----GGDAGSLAITHRAGKLIDDLKAEGRLAKLTKREAVANHANPTVDITMAEGBPV 215
QY 222 IAFGDFEITGTORYPREQIVSGLARFQGPYDLDLDDFQQALEQNGHYSVASQVADFDR 281
DB 216 APLGAVTVGKTVDCGDFIRRYSLNGGEPYSPEKLRKAADRLRLQGVFSSLTIKRAGTL 275
QY 282 LOGDRVVPKVSVEVRKHLETKGIRLDSYGLGKGIAYDYNLFNKG-----YIGSV- 333
DB 276 ARDGTITLIEVSEKHHRYFGVGAOYSTTEGLOGYWGHRNLFGQASLRLEGSVSRIA 335
QY 334 ---VWMD------KYETTLAAGISQPRNRYGNVWTSNVSNRSTTON 372
DB 336 EASSVSGMDYSAGITTKGCMFNPRFTFTSLIAKTENPDTRAKTLTGT----- 385
QY 373 LEKRAFSGGIWYVRDAGIDARLGAFLAEGRKIPGSDIDLGNHATMLTASMKROLLNN 432
DB 386 -----AGFAYELND-----TDTAAAGLEVQWADTEDAFGKNEYLTSTIPLEFVRD 430
QY 433 VLPENGHYLDGKIGTTLTGLTFLSSTALIRTSARAGYFTPEN-----KKLGT---FIIR 483
DB 431 T-----RDDKLNPTGFRASLAAPKPSYEAALNGTFFSFEGSITGYKGLGAEDRLMA 482
QY 484 GO--AGYTV-ARDNADVPGLMFRSGGASVRCGYELDSTGLAGPNSGLPRLALLVGSLE 540
DB 483 GKLSGVLVGGSLQDIPTRRFAGGGGVRGYSYQEIISPYNAAGDAGTGGRSYVVGVSVE 542
QY 541 YQLPFTRTLSGAFVHDWG-----DAAANFKRMKLKHGSLGVRWFSPPLAPFSFDIA 591
DB 543 ARKVTDTICLVFFDAGVVDGVTDFES--DIRACAGIGLRYATPFGPLRLDVA 595
```

RESULT 13

G82706

conserved hypothetical protein xfl1231 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: G82706

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: AB2515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: G82706

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-617 <SIM>

A:Cross-references: GB:AE003957; GB:AE003849; NID:g9106207; PIDN:AAF84041.1; GSPDB:GN-

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madelra, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.C.R.; de Rosa Jr., V.E.; de Sa, R.J.; Santelli, R.V.; Sawa

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: Xfl1231

Query Match 9.7%; Score 306.5; DB 2; Length 617;
Best Local Similarity 23.3%; Pred. No. 1.1e-12;
Matches 140; Conservative 89; Mismatches 232; Indels 139; Gaps 25;

```
QY 82 VLDKEQTGF---LAEAPDNVKTMLRSKGYFSKVSLSL---TEKDG-AVTVHTTPGPRTK 133
DB 79 VVKEGEGERLEVLLSQAEFOTQALEPFGYVAPTTRIDAPRONDHITVVYVDKGEVPR 138
QY 134 IANVGVAIGDILSDGNLAEEYRNALENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPL 193
DB 139 VROAHVAMTAAQAQD---HYLQRLDEDFKPLGEFNHPTYEASKVRITRRLAERGFD 194
QY 194 AKLGNTRAAVNPDPATVDLNVVDSOR-----PIAFG-DFEITGTORYPEQIVSGLARFQ 247
DB 195 ADFTHRRVEVTRAHAADIDLWESGRRYDMGPEVRHYD-----FHEGLFNPLVYWD 247
QY 248 PGTPYDLDLDDFQQALEQNGHYSVASQVADFRLQGD-RVPVKVSVTEVRKHLETKGTIR 306
DB 248 EGSYFHEGKLDRLESRLTKLDYFSSIDIQPKPEADPEGVNPVDVKLERAKSKIYTAGIS 307
QY 307 LSEYGLGKGIAYDYNLFNKGYIGSVVWMDKYETTLAAGISQPR-RNRYGNVWTSNVSY 365
DB 308 YGSESGAG-----LRAGVERRYMNARGHKMNARLDY 338
QY 366 NRSTTONLEKRAFSGGI-----WYVRDAGIDARLGAFLAEGRKIPGSDIDLG--- 414
DB 339 ---AONLKSLLTAYOIAPFKWLDGWI-----FFRAYDEOTKYIDLNRVK 381
QY 415 -----NSHAT-----MLTASMKROL-----LNNVLHPE-NGHYLDGKIGT 448
DB 382 LSAARSGQINRHILTATASLNLARGWRVYRADDGNTNVAQQSTLVYPQLEASIVDVEDAT 441
QY 449 TLGTFLSSTALIRTSARA---GYEFTPENKKIGTF-----IIRGQAGYTVARNAD 496
DB 442 FPRNGSAATVLLRGGSALGSKSNFTQLHGQLRWFHGLGASSRLILRGEAGTTWASDLVA 501
QY 497 VPSGLMFRSGGASVRCGYELDSIG---LAGPNSGLPRLALLVGSLEYQL-----PPTRTL 549
DB 502 MPPLSRFFAGGVNSIRGYAFREVPGRTPAKDPAFALGAKHVFSAAGAEFEYKGGPFP--- 557
QY 550 SGAFVHDMGDAAAANFKRMKLKH-GSGLGVRWFSPPLAPFSFDIAYG--HSDKIRWHTSLG 606
DB 558 GGAVFVDSGSA---FNRYPDWHVTGVLGRLYRSPVGPVRVDIARGNLSPSKQLYIDIG 614
```

RESULT 14

A13488

outer membrane protein BME11895 [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: A13488

R:DelVecchio, V.G.; Kapatral, R.J.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov

.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AI3488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-623 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53076.1; PID:g17983939; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11895
A:Map position: 1

Query Match 9.2%; Score 290.5; DB 2; Length 623;
Best Local Similarity 20.4%; Pred. No. 1.3e-11;
Matches 140; Conservative 106; Mismatches 248; Indels 193; Gaps 28;
QY 38 KNSPOTSEYKLPK-FPVRI-----DTQDSEIKOMVEEHLPLITQQOEVLDEKQ 87
Db 15 KKQDPDI---IDPKTYSVDVTTGDRKNADGKADLKSIVGASGLVSDADKPA--SGS 68
QY 88 TGLAEAPD--NVKTLRSGYFESSKVS-----TE--KCGAYTVHTPGP 130
Db 69 AGLAKARGDYRRILSALYGEGRYGTISIKVDCREANDIPPOTETIPNNAKVAITVDGP 128
QY 131 -----RTKIANV-----GVAIGDILSDGNLAEYRYRNALENWQOPVGS 168
Db 129 QFLSRTAISNAPPNRRDKVQTPPEAGFA-PGQEAKSGLTKLAERLAVEAWRQ---- 183
QY 169 DFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVVVDSDGRPIAFODFE 228
Db 184 -----EGYAKARVGTEDVVADHADNRVSADIALDPGRKAYTGPVS 223
QY 229 ITGTORYPEQIVSGLARFQGTGYDLDLLDFFQALEQNGHYS--GASVQADFRLQGD- 285
Db 224 VGTARMDOFVAMTGLKPGQYDPPD---DIENAKKRLGRMEVFRAMTFEEADKIEPDG 280
QY 286 RVPKVSVEVKRHKLETGRLDSEYGLGKIAVDYNNFNKGY-----IGSV-- 333
Db 281 SLPTLVNQRKPRRFGCGAEYSTIDFGVTSYWMHRNLLRGERLRFDAKVSIGGSQD 340
QY 334 -VNDMKYETTLAAGISQPNRYGN-----YWTSNVSYNRSTTQ----- 371
Db 341 NSFDPKNYTYLLGASFAKPGVYPTDVFATLDAKREVLDAYTETSINAKTGFTQIFDSE 400
QY 372 -----NLKFRASGIGWVRD--RAGIDARLGAELAEGRKIPGSDIDLGNSHATMLT 422
Db 401 LSGALYANASQGFHVDVDFKRDFTTAGLEGNL-----LYDSR----- 438
QY 423 ASMKRQLLNVLHPENGHYLDGKIGT-----TLGTFLSSTALIRTSARAGYFFTPENKKLG 478
Db 439 -----NNKPDPSGFFLVGNIIQPPYEFHYGNFATR-----FTAEGTYH 477
QY 479 TF-----IIRG--QAGYTVARNADVPGLMFRSGASSVRGYELDSIGLAPNGSVLP 530
Db 478 GFGQTORVVLGRLLKVGSIYGGSTADLPSPQLAGGSGVRGYGNICVSAGNGNIIG 537
QY 531 ERALLVGSLEVOLPFTFTLSGAVPHDMG----DAANFKRMKLKHSGLGLVRWFSPPLAF 586
Db 538 GRSLVEANGVTRITDSIGAVAFVDVAGYVGEKSFDPFSE-QMRVGVGGGLRILTSLGPI 596
QY 587 SFDIAY-----GHSDKIRWHISLQTRF 609
Db 597 RLDVAVPLNRRSGDPNPGYFVVGIGQAF 623

RESULT 15
AF2089
hypothetical protein alr2269 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AF2089
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqui
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2089
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-833 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA873968.1; PID:g17131360; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2269

Query Match 8.8%; Score 279; DB 2; Length 833;
Best Local Similarity 22.9%; Pred. No. 1.2e-10;
Matches 153; Conservative 82; Mismatches 234; Indels 200; Gaps 26;
QY 81 EVLDKQETGFLAEAPDNPVKTMLRSK-----CYFSSKYSLSLTKDG 120
Db 225 EVLVRPQSGQLTPELETQVYVNTQPGRTTTRSQLQEDINAFGTGFFSN-VQASPEDT 283
QY 121 AYTWHIT---PGP---RTKI-ANVGV-----AILGDILSDGNLAERYR 156
Db 284 PLGVRVSVFVQPNPVLKVEIQANPGTNPVSVLPQATADEIFRAQYKILNLRDLQEGIK 343
QY 157 NALLENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAK-LGNTRAAVNPDTATVDLVN- 214
Db 344 ELTKRYQD-----QGYVLNVVGAPOVSEN---GVVTIQA 376
QY 215 -----VYDSGRPIAFGDFEITGTQRYPEQIVSGLARFQGTGYDLDL 256
Db 377 EGVVENISVFRNKEGQDVNEQGP-----RCRTQDYIITRVELKPGQVFNRT 427
QY 257 LLDVQQALEONGHYSVASQVADFRLQGDVRPVKVSVEYKRHKLETGRLDSEYGLGK 316
Db 428 VQDLQVFGTGLFEDVNVSLD-PGTDPTKVVNVVNVVSSSGSIAAGAGISSSSGLFGT 486
QY 317 IAYDYNL-----FNKGYIGSVVNDMDKYETTLAAGISQPN 353
Db 487 VSYQQONLNGRNOKLGAEOVLGERELLFDLRFDPWIGG-----DPYRTSYTANIFRRS 541
QY 354 YR-----GNYWTSN-----VSYNRSTQNLKRA-----FSGGI 382
Db 542 ISLIFDKDEDIRTFDPGNPDNDGQDRPVTRLGGGVTFTRPLSANPFERAETASAGL 601
QY 383 WY-----VVRDAGIDARLGAELAEGRK-----IPGSDIDLGNHATMLTASMKRQLLNVL 434
Db 602 QYORVSTRDADGNLRKDGAVFDONGNRTSEIVPLSFSGTGEDDLLLLVLGAQRDLRNPL 661
QY 435 HPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYF-----FTPENKKLGTFIIRGOAGY 488
Db 662 QPTSGSLF--RFGVDOSVPVSGNIFLTRFRGYSQYLPVKFTGFGSKGPTIAFIQGGT 719
QY 489 TVARDNADVPGLMFRSGASSVRGYELDSIGLAPNGSVLPERALLVGSLEYQLPRTT 548
Db 720 VL---CDLPPYEFTLGGSNVRYEEGALGSG-----RSFVQASVEYRFPVFSV 766
QY 549 LSGAVFHDMDG-----AAANFKRMKLKHSGLGLVRWFSPPLAFPSFDIAYG-HSDKKIR 600
Db 767 VSGALFDVGSGLDGTSTRTAEVLNKLKSGSGYGLGVRVQSPGLPIRID--YGINDDDGDSR 824
QY 601 WHISLQTRF 609
Db 825 INFIGERF 833

RESULT 16
B87448
conserved hypothetical protein CCL1603 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-628 <STO>
A;Cross-references: GB:AE005673; NID:gl3423000; PIDN:AAK23582.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1603

Query Match	8.04;	Score 253;	DB 2;	Length 797;
Best Local Similarity	21.94;	Pred. No. 6.4e-09;		
Matches 135;	Conservative 89;	Mismatches 248;	Indels 144;	Gaps 26;
Qy	97	DNVITMLRSKGYF-----SSKVSLTEKOGAYTVHIT--PGPRTKIANVCVAILGDILSDG	149	
Db	223	EKVTFYONNGYFDRILLDTDIQTNEDKTKQTIKIVTEHGGFRMRGKVSII-----EG	274	
Qy	150	MLAEYRNAL-----NNQOPVGSDFQDQSWENSKTSVLGAVTRK-----GYPLAKG	197	
Db	275	DTNEVPKAELEKLLTMKPKGYE-----RQQTAVLGEIQNRMGSGAGAYSEI-	322	
Qy	198	NTRAAVNPDTATVDLNVVVDSGRPITAFGDFEITGTRYQRYPEQIVSGIARFPQTPPYDLDLL	257	
Db	323	SVQPLPNAETKTVDFVLHIEPGRKIYVNEIHITGNKKTREVVVRRELQWESAPYDTSKL	382	
Qy	258	LDFQALBQNGHSGASVOAEDFRLOQ--DRVPVKVSVTEVKRHKLETGIRLDSYGLGG	315	
Db	383	QRSKERVELLGYFD--NVQFQAVPLAGTPDKVDLNNLSLTERSTGSLDLSAGWVQDTGLVM	440	
Qy	316	KIADYYNLFNKGYIGSVVNDMDKYETTL-----AAGISQPNRYGNVW----	359	
Db	441	SAGVSQDMLFCGT--KSAALRASRSKTTLNGSLSTDPYFTADGVSIGYDVYTGKAFDPRK	498	
Qy	360	-TSNVSYNRSITTONLEKRAFSGGIWVRDRAGIDA-----RLGAEEFLAEGRKI	406	
Db	499	ASTSIQKYKTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHIAFDIKKYGT	558	
Qy	407	PGSDIDLGNSHATML--TASMKRQLNNVLNHPENGHYLDGKIGTTLGTPLSLTALIRTS	464	
Db	559	DGTD--GSEKGLYKGTGVGGRNKTDSALWPTRG-YLTG-----VNAEIALPGS	604	
Qy	465	RAGYFFTPENK-----KLGTFTIRGQAGYTVARDNA--DVPESGLMFRSGGASSVRGYEL	516	
Db	605	KLOYYSATHNQTNWFPEPLSKTFTLMLGVEYGIAGGYGRTKEIPFFENFYGGGLSGRGYE-	663	
Qy	517	DSIGLAGPN-----GSVLIP-----BRALLVGSLEYOLP---FTRTLGSAVFHDMDG----	559	
Db	664	--SGTLGPKVYDEYGEKISYGGNKKANVSAEELFFPMGAKDARTVRLSLFADAGSVWDGK	721	
Qy	560	-----AAANFKRM-----KLKHGSGGLGVRFSPFLAPESFDIAY-----G	593	
Db	722	TYDDNSSSATGRRVQNIYCAWNTKSTFTNELRYSAGGAVTWLSPGLPMKFSYAYPLKKK	781	
Qy	594	HSDDKIRWHIISLGRTRF	609	
Db	782	PEDEIQRFQFQFQGTTF	797	

RESULT 18
G81228
outer membrane protein Omp85 NMB0182 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81228
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Qi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: G81228
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-797 <TET>
A:A:Cross-references: GB:AE002375; GB:AE002098; NID:g7225394; PIDN:AAF40639.1; PID:g7225404
A:A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0182
C:Superfamily: protective surface antigen D-15

Query Match 8.0%; Score 253; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. NO. 6.4e-09;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

Qy	97	DNVKTLRSGYF-----SKVSLTEKDGAYTVHIT--PGPRTKIANVGVAILGDILSDG	149
Db	223	EKYTDYQNNGYDFRILDTDIOTNEBTKQTIKITHVEGGRFMKVSI-----EG	274
Qy	150	NLAEYYRNALE-----NWQOPVGSDFDQDSEWNSKTSVLGAVTRK---GYPLAKLG	197
Db	275	DTNEVPKAELEKLLTMKPKWYE-----ROOMTAVIGEIQNRMGSGAGYAYSEI-	322
Qy	198	NTRAAVNPDTATYDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGIAREPQPTPDLIDL	257
Db	323	SVQPLPAETKTVDVEFLHIEPGRKIVYNEIHITGNKNKTRDEVVRRELQWESAPYDTSKL	382
Qy	258	LDFOQALEQNGHYSGASVOADFRLOG--DRVPKVSVYEVKRHKLETCIRLIDSEYGLGG	315
Db	383	QRSKERVELLGYFD--NVQFDVYPLAGTPDKVDLNNLSLTERSTGSLDLISAGVWQDTGLVM	440
Qy	316	KIAYDYNNLENKGYIGSVVMDMKYETLL-----AAGISOPNRYGNWY----	359
Db	441	SAGVSQDNLFCTG--KSAALRASRSKTTLLNGSLSPDTPYFADGVSGLGYDVYCKAFDPKR	498
Qy	360	-TSNVSVNRRSTQNLKRAPSGGIIWVVRDAGIDA-----RLGAEFLAEGRKI	406
Db	499	ASTSIKQYKTTTAGAGIRMSVPYTEYDRVNFGLVAEHLTVNTYNNKAPKHYADFIKKYGT	558
Qy	407	PGSDIDLGNSHATML--TASWKROLNNVYHLPENGHYLDGKIGTTLGTLSSTALIRISA	464
Db	559	DGTD---GSFGKWLXGTVGWGRNKTDSALWPTRG--YLTG-----VNAEALPGS	604
Qy	465	RAGYFFTPENK-----KLGTFIIRGQAGYTVARDNA--DVPISGLMFRSGGASSVRGYEL	516
Db	605	KLOYISATHNQTWFFPLSKTFTLMLGCEVGIAGGYGRTKEIIPFENFYGGGLSGVRGE-	663
Qy	517	DSITGLACPN-----GSVLP-----ERALLVGSLEYQLP---FTRTISGCAVFHDWGD	559
Db	664	--SOTLGPVKYDEYGEKISYGGNKNKANVSAELLFPMPGAKDARTVRLSLFADAGSVWDGK	721
Qy	560	-----AAANFKM-----KLKHGSGICGVRFSPFLAPSPFDIAY----	593
Db	722	TYDNNSSATGGRVQNIYGAGNTHKSTFTTNELRYISAGGAVTWLSPLPKMKFSYAPLKKK	781
Qy	594	HSOKKTRWHISLGRF	609
Db	782	PEDEIQRFQFQLGTTF	797

RESULT 19
C70412
outer membrane protein - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #text_change 05-Nov-1999
C:Accession: C70412
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: C70412
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <AO>
A:Cross-references: GB:AE000733; NID:g2983720; PIDN:AAC07292.1; PID:g2983730; GB:AE00065
A:Experimental source: strain V5

C:Genetics:
A:Gene: Omp

Query Match 7.8%; Score 246; DB 2; Length 778;
Best Local Similarity 22.3%; Pred. No. 1.8e-08;
Matches 137; Conservative 88; Mismatches 270; Indels 120; Gaps 25;

Qy	61	DSEIKDMVEEHLPLITOOQEEVLDEKDEGTFLABEAPDNWKTM---URSKGYFSKSVSLTE	117
Db	218	DRKLGKLMKEPNI-----FLRLHAPASEETLKEDVKKIEBEFYKNEGFEVEYSYEV	271
Qy	118	KDGA----YVWHITPGPRTKIANGVAILGDILLSDGNLAEYRYNALEN---WQOPVGSDF	170
Db	272	KEEPPARKVIVKKEGPRYKLKEIKI-----EGNTLPAYSELVDNLIKKNERKGRYF	323
Qy	171	DDQSWENSKTSVLGAVTRKGYPLAKLG--NT----RAAVNPDPATVDLNNVDSGRIAPF	224
Db	324	RREVIKIKNRI----REKY--AEIGFLNTSVEERVNVPKEKYSVLLKILEGAPVYV	376
Qy	225	GDFELTGRTQYPEOIVSGLARFQPGTGYDILLDDTQOALEQN-----GHYSGASVO-	276
Db	377	KKIKIKNYETROYVIRREMRQENE-----LALKGIERTSKTRIMNLGYVEDVQIEP	429
Qy	277	-----ADFDRLOGDRVPVKYSVTEVRKHKLETGRILDSEYGLGGKLTAYDYNFLNKGYIG	331
Db	430	VPRDAWDDL-----VKIERFTGQSVGLSYNEVYTGACFTELRKGNGRGTGDI-	480
Qy	332	SVWMDMKYETTILAAGISQPRNYRGNYWTSNVSYNR-----STTQNLEKRAFSGGI---	384
Db	481	-----AGISV--SYGSLYRNNAISYTRKWKFLKKPVDLDLSAFDRRIEVDYV	524
Qy	385	VDRAGIDARLGAFFLAEGR-----KIPGSDID-----LGNSHATMLTASWK	426
Db	525	TVERTGFSVALSKSELSYWRASIGTSTORYKYSDIDPEASTYVKEQARRDSRKIFFTIT	584
Qy	427	ROLLNNYLHPENGHYLDGKIGTGTGLTGLFSLS TALIRTSARAGYEFTPENKKLGTFI-	IRGO 485
Db	585	ROTDRYLLPTKGSLEFVRNSVGVLLGGDEKEYKFEVEGAKYFSDTYPDTGIILSLKGE	644
Qy	486	AGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNG--SVLPERALLVGSLEYQIP	544
Db	645	VGFVEGYGGKKVPIDERFFYVGGDFSIGRYEYGVAGVPDNTLDPIGAKKVVITSLMTY	704
Qy	545	-FTRTLSCGVFHDWG----DAAANPKRMKLKHGSGLGVRNFPSPFAPSFSDIAY-----	GH 594
Db	705	VFKKMLYFAGFFDYGLCADKWSDFKLSNFRGGYIGVRIITPPAPITRIDWAFKTKTVPGD	764
Qy	595	SDKKIRWHISLGRTF	609
Db	765	TNRS-RIHFFVLGTFF	778

RESULT 20

AC0129
probable surface antigen YPO1052 [Imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
C:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Farnham,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague
A:Reference number: AB00001; MUID:21470413; PMID:11586360
A:Accession: AC0129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC98984.1; PID:gl5979119; GSPDB:GN00
C:Genetics:
C:Gene: YPO1052
C:Superfamily: protective surface antigen D-15

Query Match 7.8%; Score 246; DB 2; Length 795;
Best Local Similarity 21.2%; Pred. No. 1.9e-08;
Matches 142; Conservative 105; Mismatches 273; Indels 150; Gaps 30;

QY 47 VKLKPFVPRIDTQDSIKRDMVEHLPLITQOOEEVLDKREQTGLAEAP----- 96
DB 163 VDLKLVTEGVSAKIOOI-NIVGNH-----SFTYDELISRFQ---LRDEVPWNVWGDRKY 214
QY 97 -----DNVATMLRSKGF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGA 140
DB 215 QOKLAGDLETLSRFLDRGYARFNIDSTQVSLTPDKGIYVINTITEGPOK----- 267
QY 141 ILGDILLSGNLAAYRNALNMOOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKIGNTR 200
DB 268 -LNSVIVSGNLAGHQSEAEKLTKEIEGELFNGSKVTRMEDDIKKMLCRYGYAYPRV-VTQ 325
QY 201 AAVNPDTATVDLNVVDSGRPIAFGDEITGTORYPEQIVSGLARPOQPTPYDLDLLLDF 260
DB 326 PEINDDDKTVKLHNVNAGNRFYVRHIFEGNDTSKDSVLRREROMEGAWLGNQDVEAG 385
QY 261 QOALEONGHYSGASVQADFDRLQ--DRVPVKVSVTEVKRHKLETGIRLDSYGLGKIA 318
DB 386 KERLNRIGYEE--TVDETQVPGAADLVDTYKVKERNFTGSLNFGIGYGTESGVSFQVG 443
QY 319 YDYNLNFKNKGYIGSVVWMDKYEETTLAAGISOPRNYRGNYWT-----SNVSYNRSTTQ 371
DB 444 VOQDNWLGTCNTVGINCTKNDYQTYAEFTLMDP-----YFTVDGVSGLGRIFYNDFKAD 497
QY 372 NLKRAPSGGIWTVDRAGIDARLGAELAEGRKIPGSDIDLGNH-----ATML--TASWK 426
DB 498 NAD--LSG---YTNSSYGADGTLGFP-INENSL---RVGVGVVHNDLSMDLPQVAMWR 547
QY 427 RQLLNVLHPENGHYLDGKIGTTLGTLFSLTALIRTSARAGYF----- 469
DB 548 --YLESV--CERQGY-DCREGFTTDDFTLNLGWTYNNLDRGFPTSGVKSVNTKTIIVPG 602
QY 470 -----FTPENKLGTFII-RQAGYTVARNADVPVSGLMFRSGGASSYRGY 514
DB 603 SDNEFYKVTFTSAYQPLNEDRSWVLLGRGLGYDGIKSGKEMPFYENFVAGGSTVVRGF 662
QY 515 ELDSIGLAG---PNG-----SVLPERALLVGSLEYQLP-----FTRTSLGAVPHD 556
DB 663 RSNNIGPKAAYYANGATVNTSDAVGNGAMAVASIELITPTPFISEKYSNVRTSIFID 722
QY 557 MGA-----AANE-----KRMKLKHGSLGVWRFSPILAFPSFDIAYG----HSD 596
DB 723 SGPVWNTWNTAKTRAAGIPDYGKASNIRVSAGVALQWMSPLGLVFTYAKPVKDYEGD 782
QY 597 KIRWHISLG 606
DB 783 KSBQFQFNIG 792

RESULT 21

B82099
surface antigen VC2252 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: B82035; MUID:20406833; PMID:10952301
A:Accession: B82099
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HEI>
A:Cross-references: GB:AE004237; GB:AE003852; MID:99656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2252

A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 7.2%; Score 227; DB 2; Length 803;
Best Local Similarity 21.7%; Pred. No. 3.6e-07;
Matches 135; Conservative 95; Mismatches 225; Indels 168; Gaps 31;

QY 84 DKPOTGLAEAPDNVKTMLRSKGYFSSKVSILT-----EKDGAY-TVHITPGPRTKIAN 136
DB 212 DKYKKQVLAGDI-EALRTYVLDRLGKLFQVDSQVLAISDPKKGVIITLNLNEGEPTVSK 270
QY 137 VGVAILGDILLSGNLAAYRNALNMOOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKL 196
DB 271 --VQFRGELM--GKEAEF--TSLIPFE--IGETVNGSAVTRLEESVKVLGSGYAYPOV 322
QY 197 GNTAAVNPDTATVDLNVVDSGRPIAFGDEITGTORYPEQIV----- 240
DB 323 -RTIPEFDEKQOVSLLVHVHVEAGKRVYVRDIRFVGNNTSDEVLRREROMEGSWLNSKD 381
QY 241 --SGLARFO-----PGTPYDLILLDFQOALEON-----GHYSGASVQ 276
DB 382 IETGKTRLNRIGYEEVQTVRVPQSEDOVDLVYSVKEANSNGVNFVGCGYGTESGVSFQ 441
QY 277 ADPDR----LQGRVPVKVSVTEVKRHKLETGIRLDSY-----GLGKIAVDYY 322
DB 442 VGLQDNFLGSGNRGVGNAMINDYQKN-----LTLEYRDPYWNLDVSGVSGKV---FY 491
QY 323 NLFNKGYIGSVVWMDKYEETTLAAGISOPRNYRGNYWTNVSYNRSTTQNLKRAPSGGI 382
DB 492 NQPEASEAGIVDTNDSYGTSLTWGF--PFD-----ELNRFEEGI 529
QY 383 WYVRDRAG-IDARLGAE-FLAEGRKIPGSDIDLGNHAT---MLTASWKRLNNVLHPE 437
DB 530 GYTHNKIGNLTPVLOVENFLA---AQASNIDSGNLLTDDFDNLNLSWTRNLSNFPPT 585
QY 438 NGHYLDGKIGTTLGTLFSLTALIRTSARAGYFTTPENKLGTFIIRQOAGY-----TVA 491
DB 586 AGNHORAFYKMTVPG--SDAQYFKLQYDVRYFPLTKKHEFTLLRLGRGLGNGYGTODG 643
QY 492 RDNADVPVSGLMFRSGGASSYRGYELDSIGLAGP-----NGS-----VLPERAL 534
DB 644 KDNL-PFFYENFYAGGFTSLRGFSNS---AGPKAVYRDYSGSNNGSDTATDSDSVGNAI 699
QY 535 LVGSLEYQLP-----FTRTSLGAVPHDMG---DAAANFKRMKLKHGS----- 573
DB 700 ALASVELIVTPPASEARNQIRTSIFYDMASVWDTEFDY-RGKADYGNQYYDYSDPTN 758
QY 574 -----GLGVRWFSPPLAFPSFDIA 591
DB 759 YRSSYGVALQWSPMGPLVFSLA 781

RESULT 22

AB0530
outer membrane protein precursor yaeF [imported] - Salmonella enterica subsp. enteritidis
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0530
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church-
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far-
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:g16501505; GSPDB:GN00176
C:Genetics:
A:Gene: yaeF

C:Superfamily: protective surface antigen D-15

Query Match 7.1%; Score 226; DB 2; Length 803;
Best Local Similarity 20.1%; Pred. No. 4.2e-07;
Matches 136; Conservative 102; Mismatches 284; Indels 154; Gaps 25;

QY 47 VLKPKFPVIRDTQSEIKDMVEEHLPLITQOOEVLKQEQTFIAEAP----- 96
DB 163 VDLKLVFQEGVSAKIQI-NIVGNH-----AFSTEELISHFO---LRDEVPMVNVVGRDKY 214
QY 97 -----DNVKTMLRSKGTF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGA 140
DB 215 QKQKLAGLETLRYLLDRGFARFNIDSTQVSLTPDKKGIITVINITEGDQTKLSGVQVS 274
QY 141 ILGDITLSDGNLAERYNALENMQQVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTR 200
DB 275 -----GNLAGHSAEIEKLTKEPGEIYNGTKVTKMEDDIKLLGRGYAYPRV-QSQ 325
QY 201 AAVNPDTATVLDNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPTDLDLLDF 260
DB 326 PEINDADKTVLRNVNDAGNRFYVRKIRFEGNDTSKDSVLRREMRQMEGAWLGSDLVDQG 385
QY 261 QOALPONGHYSGASVOADFRLQ--DRVPVKVSVTEVKRHKLETRILDSFYLGGRIA 318
DB 386 KERLNLRLGFFE--YVDTQTVPGSPQDVVYVKERNTGSFNGFYGTGSGVFSQAG 443
QY 319 YDYNLFNKGYISYVWDMKYETTLAAGISOPRNYRGNYWT----- 360
DB 444 VQDNWLGTYSVGNGTKNDYQTYSELSTNP-----YFTVGVSLGGRIFYNDPEAD 497
QY 361 -----SNVSYNRSTTNL---EKRAFSGGIWYVRDR-AGIDARLGAETFAEGRKIPGS 409
DB 498 DADLSDYTNKSYGTDVTLGFPINNEYNTLRAGLGYVHNSLNQMQPQIAMDRYLE----- 550
QY 410 DIDLGNHATMLTA-----SMKROLLNNVLPENGH--YLDGKIGTTLGTLFSLSTALI 460
DB 551 --SMGDDPASDFADDTFTNGTWNKLDRGYFTDGSRVNLTGKV-TIPG---SDNEY 604
QY 461 RTSARAGYFTPENKKGITFIIRGAGYTVARNADVPGLMFRSGGASSVRGYELDSIG 520
DB 605 KVSIDTATYVPIIDNHKVVLRGTRWGYDGLGGKEMPFYENFVAGSGSTVRGQSNTIG 664
QY 521 --LAGPNS-----VLPERALLVGSLEYQLP-----FTRT 548
DB 665 PKAVYKNGAHTSWDDDDYEDCTQESGCKSDDAVGGNMAVASLEFIPTPFISEKYANS 724
QY 549 LSGAVFHDWG-----DAANFKRMKLKHGSLGGLVGRWFSPLAPSFDAIYG- 593
DB 725 VRTSFFDMGTWDTNWDPPSAPSDVPDYSDPGNIRMSAGIALQMSPLGLPLVFSYAQP 784

RESULT 23

A64742
hypothetical protein b0177 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64742
R:Blattner, F.N.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64742
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-810 <BLAT>
A:Cross-references: GB:AE000127; GB:U00096; NID:g1786370; PIDN:AAC73288.1; PID:g1786374;
A:Experimental source: strain K-12, substrain M61655
C:Superfamily: protective surface antigen D-15

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;

QY 97 DNVKTMLRSKGTF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGVAILGSDIG 149
DB 224 ETLASYYILDRGYARFNIDSTQVSLTPDKKGIYVTVNITEGQYKLSGVEVS-----G 275
QY 150 NLAERYNALENMQQVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTAT 209
DB 276 NLAGHSAEIEKLTKEPGEIYNGTKVTKMEDDIKLLGRGYAYPRV-QSMPEINDADKT 334
QY 210 VLDNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPTDLDLLDFQQALEQNGH 269
DB 335 VKLRNVNDAGNRFYVRKIRFEGNDTSKDAVLRREMRQMEGAWLGSDLVDQGERLNLRLG 394
QY 270 YSGASVOADFRLQ--DRVPVKVSVTEVKRHKLETRILDSFYLGGKIAIYDYNLFNK 327
DB 395 FE--TVDTQTVPGSPQDVVYVKERNTGSFNGFYGTGSGVFSQAGVQDNLWLTG 452
QY 328 GYISVWDMKYETTLAAGISOPRNYRGNYWT-----SN 362
DB 453 GYAVGINTKNDYQTYSELSTNP-----YFTVGVSLGGRIFYNDPEADLSDYTN 506
QY 363 VSYNRSTTNL---EKRAFSGGIWYVRDR-AGIDARLGA-EFLAEGRKIPGSDIDLGNH 417
DB 507 KSYGTDVTLGFPINNEYNSLRAGLGYVHNSLNQMQPQVAMWRYLYSMGEHP-STSDODNSF 565
QY 418 AT---MLTASWKROLLNNVLPENGH--YLDGKI-----GTTLGTFLSSTALIRT 462
DB 566 KTDDFTFYNGTYNKLDRGYFTDGSRVNLTGKVTIPGSDNEYKYVTLDT----- 615
QY 463 SARAGYFTPENKKGITFIIRGQA--GYTVARNADVPGLMFRSGGASSVRGYELDSIG 520
DB 616 ---ATYVPIDDDHK---WVILGRWGYDGLGGKEMPFYENFVAGSGSTVRGQSNTI- 668
QY 521 LAGPNSVLPER-----ALLVGSLEYQLP----- 544
DB 669 --GPKAVYFPHQASNYDPDYECATQDAGKDLCKSDDAVGGNMAVASLEFIPTPFIS 726
QY 545 --FTRTSLGAVFHDWGDA-AANFKRMK-----LKHGSLGVRWFSPLAPSF 589
DB 727 DKYANSVTSFTSFDWGTWDTNWDSDSYSGYSDPSNIRMSAGIALQMSPLGLPLVFS 786
QY 590 IAYG---HSDKKIRWHISLG 606
DB 787 YAQPFKKYDGDKAQEQFQFNIG 807

RESULT 24

C90651
hypothetical protein ECs0179 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90651
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90651
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833602.1; PID:g13359635; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs0179
C:Superfamily: protective surface antigen D-15

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;

Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;

[illegible]

RESULT 25
C85502
hypothetical protein yaeT [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85502
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85502
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <STO>
A:Cross-references: GB:AE005174; NID:gi2512902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:Z01
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yaeT
C:Superfamily: protective surface antigen D-15

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;
Qy 97 DNVTMLRSKGYF-----SSKSVLT-EKOGAY-TVVHTTPTGRTKIANGVAILGDIISDG 149

Db	224	ETLSRYLDRGYAREFINDSTQVSLTDPDKGIYTVNITEGDOVKLGSVEYS-----G	275
Qy	150	NLAEYRNALNEMOQPGVSDFDODSWENSKTSVLGAVTRKGYPKLKIGNTRAARNPDAT	209
Db	276	NLAGHSABIEBOLTKEIEGELYNGTKVTKMEDDTKLLGRYGYAYPRV-QSMFEINDADKT	334
Qy	210	VDLNVVDSGRPIAFGDEITGTQRYPEQIVSGLARPQPGTXYDLLLLDFOOALEQNH	269
Db	335	VKLRVNDAGNRFRVYRKIRPEGNDTSKDAVLRREMRQMEGAWLGSULLDQGGKRLNRLG	394
Qy	270	YSGASVOADFORLOG--DRYPVKYSVTEVRHKLETCIRLDSEYGLGGKIAYDYNNLPNK	327
Db	395	FE--TVTDITQVRPGSPQDVVYVKVERNTGSEFNFGIYGTGSEGSVFGAQVOODNMLGT	452
Qy	328	GYIGSVWMDMDKYETTLLAAGISQSPRNYRGNYWT-----SN	362
Db	453	GYAVGINTKNDYQTYAELSVTNP-----YFTVDGVSGLGRLFYNDFOADDADLSDTYN	506
Qy	363	VSYNRSTQNL--EKKRAFSGIHWYVDR-AGIDARLGA-EFLAERKTKPGSIDLGNSH	417
Db	507	KSYGTDVTLGPPINEYNSLRAGIYVHNSLNQPOQVAMWRYLYSMEHIP-STSDQDNPS	565
Qy	418	AT--MLTASMKROLLNNLHPENGH--YLDGKI-----GTTTGLTFLSSTALIRT	462
Db	566	KTDDEFNYGTYNKLDRGYFPTDSRVNLTKVFTPGSDNEYKKYVLDT-----	615
Qy	463	SARAGYFTTPENKKLGTFIINGQA--GYTVARNADVPVSGLMFRSGASSVRGYELDSIG	520
Db	616	---ATYVPIDDDHK--WVVLGRTRWYGDGLGCKEMPFYENFYAGCSSTVRFGQSNTI-	668
Qy	521	LAPNGSVLPER-----ALLVGSLEYOLP-----	544
Db	669	--GPKAVYFPQASNYDDYDECATODGAKDLCKSDDDAVGGNAMAVASLEFITPTPFIS	726
Qy	545	---FTRLSSGYFHDWGDA-AANEXRKM-----LKHGSLGVRWFSPLAPSFSD	589
Db	727	DKYANSVRTSFFWDGTVTWDTNWDSQYSGVPDYSDFSNIRMSAGITALQWMSPLGLVFS	786
Qy	590	TAYG-----HSDKKIRWHISLG	606
Db	787	YAQPFKPYDGDKAEGQFQFNIG	807
		RESULT	26

RESULT 26
H83190
probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain P
H83190)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83190
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic P
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83190
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <STO>
A:Cross-references: GB:AE004784; GB:AE00491; NID:g9949799; PIDN:AAG07036.1; GSPDB:G
A:Experimental source: strain PA01

A.Gene: PA3648
C.Superfamily: protective surface antigen D-15

	Query Match	6.8%	Score 215.5;	DB 2;	Length 797;
	Best Local Similarity	20.0%;	Pred. No. 2.le-06;		
	Matches 133; Conservative	83;	Mismatches 240;	Indels 209;	Gaps

QY 105 SKGYESSKVSUTEKDGYVTHITGPRTKA-----VAALGD-ILSDGNL 151
:: :: :: :: ::
:: :: :: :: ::

DG 142 AGGRYSAGIN-----AEVIOPNRNVALKINIEGTVAALSHLVNVTGVFEEDL 192
:: :: :: :: ::


```
Db 233 FDFRVISSVAELAPDKNGFAVYTTLEEGPKYKPGKITVETELKKL-DGNLQAQI----- 285
QY 162 WOOPV--GSDFDODSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDG 219
Db 286 --LPVRTGQLYEDERIEQATDALTFAAGAAGFAVDV-PRPVVPRETXYVDVVOVRG 342
QY 220 RPIAGDFEITGQRYPEQIVSGLARFQPTPYDLDDLLDFQOALEQNGHYSGASVQADF 279
Db 343 PRVYVDRIDVGNTRPLDVLRLRELEVAEGDAYNRVLVDRSKNNMRLGFFKEVEIE-DA 401
QY 280 DRLOGDRVPVKSVTEVKRHKLETGIRLSEYGLGKIAIYDYNLFNKYIGSVVWMDK 339
Db 402 PGSAPDRTSLRVKVEQPTGELS-----FSAQYS-----SIDK 434
QY 340 YETTLAAGISOPRNYGNGWTSNVSYNRSTTONLEKRAFSGGIWYVRDAGIDAR-LG-- 396
Db 435 --LVLDVGITE-RNFGR-----GQNLRARASVGLRQOIDFGFSEPRELGRN 479
QY 397 -----AEFLAEGRRKIPGSDIDLG-----NSHATMLTA 423
Db 480 LVAGVNLTYFRYDLSEFAAYDTKSVGGDVRFGLPTLNDSSMSLRYTVRQDEVSVADSLCA 539
QY 424 SHKROLLNVVLPENCHYLDGKIG-----TTLCTPL-----STALI 460
Db 540 SGS---VSQILCQAGAYITSLIGYGLRIDKRNDRPNPTRGNFADLNQDLAGVGGDKYL 596
QY 461 RTSARAGYF--FTPENKILGTFTIRGOAGVTVARDNADVPGLMFRSGGASSVRGYELDS 518
Db 597 KTEADAGWNGFT--KDL-VFSATGSGFYIEGWGDNVWRINDRFRGG-TSRGFEIAG 651
QY 519 IGLAGPNSVLPERALLVGSLEYQ-----LPFTRTLSGAVFHDMDGAA----- 561
Db 652 IGRPDISSFNSMGAKLYAISTELTVPTPLPEQYGIKAALFSDVGTAGLLDDVDQRSP 711
QY 562 ----ANFK-RMKLKHSGSLGVRFSPFAPSFDAIYCHS-----DKKIRWHISLGRF 609
Db 712 GVFDPNIKDNLGRASAGISIDKSPMGPIRFDISRLSKEDYDRTETFRFSTSTRF 768

RESULT 29
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C:Accession: JC4078
R:Black, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus inf
A:Reference number: JC4078; MUID:95255676; PMID:7737523
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <PLA>
A:Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:9537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1/19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.6%; Score 209.5; DB 2; Length 797;
Best Local Similarity 20.0%; Pred. No. 5.3e-06;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEHLPLITQQQEEVLDEKQTFGLAEAPDNVKTMLRSKGYFSKVSALT----- 116
Db 184 ESVSSSTLQOMELQPDSSWKLWGNKFEQAQFEKDLQSIIRDYVLYNNGYAKAQITKTDVQL 243
QY 117 --EKDGA-YVHTITPGPRTK-----IANYG--VAIIGDILSDGNLAEEYRNALENWQOP 165
Db 244 NDEKTKVNVITDNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFTRS----- 296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNP-----TATVDLNVVVDGRP 221
```

```
Db 297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGQRYPEQIVSGLARFQPTPYDLDDLLDFQOALEQNGHYSGASVQADF 281
Db 344 LVTRQLRFFGNVTVSADSTLQEMRQOEGTWYNSQLVELKIRIDRTGTFFE--TVENRIDP 401
QY 282 LQO--DRVPVKSVTEVKRHKLETGIRLSEYGLGKIAIYDYNLFNKYIGSVVWMDK 339
Db 402 INGSNDVDVYVKERNTGSIINFYGYGTESGISYQASVKQDNPLGTGAASVIACTKND 461
QY 340 YETTLAAGISOPR-----NYRGNWTSNVSYNRSTTONLEKRAFSGGI--WYVRDRA 389
Db 462 YGTSVNLGYTEPYFTKDGVSGLGNVEFNYSKSDTSNYSKRTYTGSNVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----CRKIPGSDIDLGNSHATMLTASWKRQLL 430
Db 522 SYVYVGLHYNTKISNFALEYNRNLYTQSMKFKNGIKTNDFF-----FSGWYNSL 573
QY 431 NVLHPENGHYLDGKITGTTLGTPLSTALIRTSARAGYFTPENKILGTFTIRGOAGYTV 490
Db 574 NRCYFTKG--VKASLGGRTVTPGSDNKYYKLSADYQGVPLDRDLHWVSASAKSAGYAN 631
QY 491 ARNADVPGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531
Db 632 GFGNKLPLPYQTYTAGGIGSLRGFAFGSI---GPNAIYAEYGNNGSGTGTFFKISSDVIG 688
QY 532 RALLVGSLEVLDP-----FTRTLSCAVPHDM-----GDAANFKM----- 567
Db 689 NATATASAEILVTPPVSKSNTVTSLSFVDAASVNTWKWSKXNGLESVDVLKLPDYG 748
QY 568 ---KLKHGSLGVRFSPFAPSFDAI-----YGHSDKKIRWHISLGRF 609
Db 749 KSSRIARASTGVQWQSPITGLPVFSYAKPIKKYENDDVE-QFOFSIGGSF 797

RESULT 30
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C:Accession: F64102
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidma
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-808 <TIGR>
A:Cross-references: GB:I42023; TIGR:HI0917
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen

Query Match 6.6%; Score 209.5; DB 2; Length 808;
Best Local Similarity 20.0%; Pred. No. 5.4e-06;
Matches 130; Conservative 103; Mismatches 277; Indels 141; Gaps 24;

QY 61 DSEIKDMVEHLPLITQQQEEVLDEKQTFGLAEAPDNVKTMLRSKGYFSKVSALT----- 116
Db 197 ESVSSSTLQOMELQPDSSWKLWGNKFEQAQFEKDLQSIIRDYVLYNNGYAKAQITKTDVQL 256
QY 117 --EKDGA-YVHTITPGPRTK-----IANYG--VAIIGDILSDGNLAEEYRNALENWQOP 165
Db 257 NDEKTKVNVITDNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFTRS----- 309
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNP-----TATVDLNVVVDGRP 221
Db 310 ----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 356
```

Qy 222 IAFGDFELTGRTYRPEIQISGLARFPQGTPDYDLDLLDFQQAALEQNGHYSYGASVQAADFDR 288
:
Db 357 LTVRLRFEGNTVSADSTLRQMQRQBGGTWNYSQLVELGRILRDRTGFEE--TVENRIDP 414
:
Qy 282 LOG--DRVPPKVSYTEVKRIHKLFTGLDSEYLGGCKIAYDIYNLENPKNGYCISGVWDMDK 339
:
Db 415 INGSNDEVDVVYKVKERNTGSIINFGIGYTGESISYAQVKODNFLGTGAAYSIAGTKND 474
:
Qy 340 YETTLAAGISOPR-----NVRGNYMITSNNYSNRS--TTONLEKRAPSGGI---WVRDRRA 389
:
Db 475 YGTSVNLYGYTEPTFKDGVSILGNVFENNDSKSDTSSNKYKTYYGSNVTLGFPVENNN 534
:
Qy 390 GIDARLQ-----AEFLAE-----GRKPISGDIDLGNSHATMLTASMKRQLL 430
::
Db 535 SYIVVLGHTYNTKISNFALERNRNLIQTSMKFKGNGIKTNDFD-----FSPGMNNSL 586
:
Qy 431 NNVLHPENGHYLOGKIOTTLGTLSSTALIRTSARAGYEFTPENKKLTFFIIIRGOAGTYV 490
:
Db 587 NRGVPTPKG--VKASLGGRRVITPGSONKYVKLSADVQGEFYPLDRDLHWLVWSAKASAGYAN 644
:
Qy 491 ARDNADVPGLMFPSGCCASSRVGYELDSLCLAGPN-----GSVLPERRA 533
:
Db 645 GFGNKRLPFYQTYTAGSIGLRFAGYCSI---GPNAIYAEHGNGNGTFKKISSDISDIGNA 701
:
Qy 534 LLVGSLSEYQLP-----FTRTSLSGAVFHMDMGDAANF----- 564
:
Db 702 ITTASAELIVPTFEVSDKSKONTVRTSLF---VDAASVWNKWKSKSGLDNWLKSLDPY 758
:
Qy 565 -RKMKKHGSLGVRYMPSPFIAPESFDITA-----YGHSDKKIRWHISLGTFRF 609
:
Db 759 GKSSRIRASTGCVGFMQSPIGPLVFSYAKPIKKYENDDVVE-QQFPSIGSFP 808
:

RESULT 31

A81430

C:outer membrane protein Cj0129c [Imported] - Campylobacter jejuni (strain NC

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2000

R:Cathillon, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham

C.W.; Quail, M.; Rayandream, M.A.; Rutherford, K.M.; Vanvliet, A.; Whitehead

Nature 403, 665-668, 2000

A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni

A:Reference number: AB1250; UID:20150912; PMID:10688204

A:Accession: AB1430

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-739 <PAR>

A:Cross-references: GB:AL139074; GB:AL111168; NID:g9697505; PIDN:CAB72613.1

A:Experimental source: serotype O2, strain NCCTC 11168

C:Genetics:

A:Gene: Cj0129c

Query Match	6.5%	Score 206;	DB 2;	Length 739;
Best Local Similarity	18.3%;	Pred. No. 8.1e-06;		
Matches 140;	Conservative 102;	Mismatches 312;	Indels 210;	Gaps 24;

Qy	19	AYAPAADLSENKAAGFALFNKSP-DTESVKLAKPKP-PVRIDT-----	59
Db	13	AMANAATIKDIFKLTIGLNHLNLSNTSAINIAGLKIGEENIPAKINTAILNLNKQYFENIAVE	72
Qy	60	QDSEIKDMVEEHLPLIT-----OOQEEVLDDKEQTGLAE---EAPDNVKTML	103
Db	73	NNNGILEIIVTEPTAKVTITGASNDRKQVESILGIRKGTLLDGNLKEAETIKAVY	132
Qy	104	RSKGYSFSSVSLTEKDGATVYHTTTPGPRTK-IANVGVAIGDITSLDGNLAETYYRNALENW	162
Db	133	EAKSYFDTIVEYKKK----TLENTDGLLEFFIVNRGNIIDNVHLSGAKKFSYSDIEPA	188
Qy	163	QQPVGSDPDDQSW-----ENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATV	210
Db	189	VVNKEFEFGMGMGRNDGLKLVFELNSDSSRIADHEYMKKGYLQDVQSPYKLTYYTDTQYA	248

QY	211	DLNVVVDSGRPIAGDFEITG-----	231	
		: : : : :		
Db	249	NLTVPFKGKPKIKSISIENPLFDDKQNAQTVKDLRSSAGKTIINEDIRKDVKTITQS	308	
QY	232	-----TQRYPE-----QIVSGLAR-----	247	
		: : : : : : :		
Db	309	ADLGAVFVEVYVDIOKNDOTQEAIVVFKVIPHDKVYIRNVIISGNSRTVDVRVIRRELYIT	368	
QY	248	PCTPYDLDLLLOFOALAEONHSGASVQADFRLQGDVRPVKVSVEYKRHKLEGIIRL	307	
Db	369	EGNLYNRDTLSEKNALKRTSYFDDVNIKE--EKVDDTHDLIDVYKEASTGAISGIGY	426	
QY	308	DSEYGLGCKIADYNNLFNKGYIGSVVMDMDKYETTLLAAGIS--QPRNRYGNWTWTSNVSY	365	
		: : : : : : : : : :		
Db	427	GSSDILLNASLSDTNIFGSGIKSSV--SVDKSDDTLSGRISLVNPRVLDSQYSIGSTLY	484	
QY	366	NRSITONLEKRAFSGGIIW--YVRDRAGIDARLGAEP-----LAEGRKIPGSDI-----	411	
Db	485	SDNYE-----WDNYSEKNYGFDTIGRQFARYYVNSULTYNLEQSDIYHLSPTL	532	
		: : : : : : : : : :		
QY	412	-----DLGNSHATMLTASWKRRQLLNVLHPENGHYLDGKIGTTLGFLSSTALIRTSARA	466	
		: : : : : : : : : :		
Db	533	LRTGYELGKSKTSKSIITPAITFNDTDYYLPRSG-----IIASTSLEYAGLGG	579	
QY	467	GYFTFPPENKKLGTF-----IIRGQAGYTVARDNADVPGLMFRSGGASSVRYGYE	515	
Db	580	DOEFISSSKFNFOGLQYIGYDLIVRYKASFYKWDECYLPINQRIYLGIRSGIRSGFE	639	
QY	516	LDSGLAGNGSVLPERALLVGSLEYOLPSTR--TLSGAVPHDMGDAAA-----NFKRMKL	569	
		: : : : : : : : : :		
Db	640	SRTVSPKNQWDEIGGTAFANSVELSFPLIDRIKLRGVSFFDYGMIGRKNLDEIKRM--	697	
QY	570	KHGSGLGVRFSPFAPSFDAIYCHSDKK----IRWHIISLGTRF	609	
		: : : : : : : : : :		
Db	698	--STGIEWITPIGPIQLVFAKPLNDKKGDDNSFEFNIGTRF	739	
		: : : : : : : : : :		
RESULT	32			
S77409				
		hypothetical protein - Synecocystis sp. (strain PCC 6803)		
A:Species:	Synecocystis sp.			
A:Variety:	PCC 6803			
C:Date:	25-Apr-1997	#sequence_revision	25-Apr-1997	
C:Accession:	S77409	#text_change	08-Oct-1999	
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yama				
DNA Res. 3, 109-136, 1996				
A:Title:	Sequence analysis of the genome of the unicellular cyanobacterium s.			
A:Reference number:	S74322; MUID:97061201; PMID:8905231			
A:Accession:	S77409			
A:Status:	nucleic acid sequence not shown; translation not shown			
A:Molecule type:	DNA			
A:Residues:	1-861 <RAN>			
A:Cross-references:	EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAAL17512.L			
A:Note:	the nucleotide sequence was submitted to the EMBL Data Library, Junc			
C:Genetics:				
A:Start codon:	GTG			
Query Match	6.4%	Score	202.5; DB 2; Length	861;
Best Local Similarity	20.4%	Pred. No.	1.8e-05;	
Matches	131; Conservative	77; Mismatches	254; Indels	179; Gaps
QY	76	TCOOEEVLDEKOTGFLAEEAPDNVKTMLRSKGYFFSKSVSLTEKDGAYTVHITPGPRTKIA	135	
		: : : : : : : : : :		
Db	293	TOLQEDV-----NAIATGYFSN-VRVAPSDTPIGVRVTVFQVQANPV	333	
QY	136	NVGV-----AIIIGDILSDGNLAEYYRNALENKQOPVGSDFDQDQSWENSKTSV	182	
		: : : : : : : : : :		
Db	334	FTGLNIRTVETAEGKEIRILPOEVVDETFGEQYGKILNLRELQEGIKTTNEWYSN-----	388	
QY	183	LGAVTRKGYPLAK-LGNTRAAVNPDTATVDLVNVVDSGRPIAFGDFEITGTQRYPEQIVS	241	
		: : : : : : : : : : : : :		

Query Match 6.4%; Score 202.5; DB 2; Length 861;
Best Local Similarity 20.4%; Pred. NO. 1.8e-05;
Matches 131; Conservative 77; Mismatches 254; Indels 179

[illegible]

Db 365 GNTHKHDVILHETSLPPGDTFORLKLDETPETRLNMGYFKSVSVYVTRSQDLPLDLSNNL 421
 Qy 287 ---VPRKVSVTEV-----KHKLETG---IRLQSE 310
 Db 425 YRDVFIEVETETCNGLGLFGPSSIDHLPFGAEISESNFDLCARHLFKKFKSLRGGE 484
 Qy 311 Y-----GIGGKIADYNNLKNKYIGSVW-----DMDKYETTLAAGISOPRNVRYGNYWTS 361
 Db 485 YLFKLANLGGDKVT-DYVKKVTKPHFLNTPMILGVELDK---SINKALSK-----DYSVD 534
 Qy 362 NVSYNRSTTONLEKRAPSGGIGWVRDPAIGDARLGAFLAEGRIKPGSIDDLGNSHATML 421
 Db 535 TYGNIISTITLNDK-LKYGMVY-----RGSQTSLSLKKTENPNKPGPLD-----SNKGV 586
 Qy 422 TASWKROLLNNVLHPEN-----GHYLDGKIGTTLGTFLLSSTALIRTSARAGYVFFRP 472
 Db 587 SAAGLNVLSDISDNPRKPTWCIRSSLNFLSGLGTYQYTKLTA-----SGSIYRL 637
 Qy 473 ENKKLGTFTIRGQA-----GYTVARNADVPGLMFRSGGASSVRGYELDSTG-----LA 522
 Db 638 LTKK-GVLKIRGEAKFKTPGGTTAQ---GIPVSERFFLGGESTVRGYKPFIIIGPKFSPT 693
 Qy 523 GPNGSVLPERRALLVGSLEYOLPFTRLTSGAVFHDMDAAANFRMKLKH---GSGLGVR- 578
 Db 694 EPOGGL--SSLLTTEEPQYPLISOPSIANAFLVDSFGICTEYTIIRLKNLCSAGFGLR 751
 Qy 579 -----WFSPLAP 585
 Db 752 DMNNVPMILGWGHPFRP 769
 RESULT 34
 E82731
 outer membrane antigen XF1046 [imported] - Xylella fastidiosa (strain 9a5c
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: E82731
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nuc
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82731
 A:Status: preliminary
 A: Molecule type: DNA
 A:Residues: 1-784 <STM>
 A:Cross-references: GB:AE003941; GB:AE003849; NID:g9105978; PIDN:AAF83856.
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Al
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D
 as-Neto, E.; Docena, C.; El-Dorry, H.; Faciniani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franc
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama
 chado, M.A.; Madeira, A.M.B.F.; Madeira, H.M.F.; Marino, C.L.; Marques, M.
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.;
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.;
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli,
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.
 M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Ve
 A:Reference number: A59328
 A:Reference number: A59328
 C:Genetics: annotation
 A:Gene: XF1046
 C:Superfamily: protective surface antigen D-15
 Query Match 6.2%; Score 197.5; DB 2; Length 784;
 Best Local Similarity 20.5%; Pred. No. 3.3e-05;
 Matches 136; Conservative 105; Mismatches 291; Indels 131; Gaps
 Qy 36 LFKNKPDPSTSVKLPKPPVR-----IDTQDSEIKDMV-----BEHLPLITQQOEEVLDR 86
 Db 164 LDRNRVDVYIIKGGKAAKTRHNLIGTEKFNKNKDVMSWESKEHNWASWYRRDDOYSKE 223

Matches 121; Conservative 87; Mismatches 272; Indels 98; Gaps 25;

```
Qy 93 EAPDNVKTMLRSKGY-----FSSKVSILTEKDGAYTVHIT--PGPRTKIANVGVAILGDI 145
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 234 EEA---LRQFYNGRYADFRVSSDAVLDESKEVYITITVDEGKKYDFGNVAV---ES 286
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 146 LSDGNLAEEYRNALENMQOPVGSDFDQDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVP 205
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 287 TVPGVDGSELOGLVETRO---GASYSAKEVQOSMEAIKSRVAGEGYPPARV-TPRGDRDM 342
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 206 DTATVDLNVVDSGRPTAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDFOQALE 265
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 343 SGNTIGVTYIVDQGERAYVERIEIRGNTRDRDYVIRREFDISEGDAFNQTIITAARLRE 402
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 266 QNGHYSASVQADFDRLQG-----DRVPVKVSVTEVKRHKLETGRLDSEYGLGKIAYDY 321
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 343 SGNTIGVTYIVDQGERAYVERIEIRGNTRDRDYVIRREFDISEGDAFNQTIITAARLRE 402
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 266 QNGHYSASVQADFDRLQG-----DRVPVKVSVTEVKRHKLETGRLDSEYGLGKIAYDY 321
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 403 ALGYFSKVNIST-----AGGSAPDRVIVVDVEQSTGSGFAGAGYSONDQVLLSEASVEE 457
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 322 YNLFNKG-YIGSVVMDMKYETTLAAGISQPRNRYGNWTSN-VSYNRSTTQNLKRAFS 379
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 458 KNFLGRGOYI-----RVAAGAGEDDARTYSLSTPEPYFLGYRLAAGDFDLFKNOSK 507
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 380 GGIWVDRDRAGIDARLGA-----EFLAEGRKIPGSDIDLGNHSHATMLTA- 423
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 508 SEDYNYDEQGFALRVTAIPITENLSTTFKYTKQINYEKGQDQWQNNANLAEPYQALIRGE 567
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 424 SWKROLLNNVLPENHGYLDGK-----IGTTLGTFLLSSTALIRTSARAGYEFT-- 471
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 568 DWTOSILSNTL---NYNLTDDRNMPREGWQAALTFNEFAGLGDSSEYKYIYAKARYYYTLS 624
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 472 PENKKLGTFIIRGOAGYTV-ARDNADVPGLMFRSGGASSVRYGELDSIGLAGPNGSVLP 530
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 625 DEYDVIGS--LTGQAGHVMPGDNLLVPDQKFE---GGRQVRGFKNDGIGPRIGSDSI-G 678
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 531 ERALLVGSLEYQLP-----FTRTLGSAVFHDMGDAAN----FKRMK-----LKHGSGL 575
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 679 GTTYFAASAETAPMPGVPEDFGLRLAG--FVDAGTMYGNKVSTSQTVKDDNSIRASAGI 736
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 576 GVRWFSPLAPSFDF-----IAYGHSDDKIRWHISLGTRF 609
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 737 GVMNASPPGPIRVDAIPIAKEDYDEEQRFRFGMSNTF 774
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

```
RESULT 37
AE2746
group 1 outer membrane protein precursor ompl [imported] - Agrobacterium tumefaciens (st
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE2746
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42387.1; PID:g17737996; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Map position: circular chromosome
```

Query Match 5.8%; Score 185; DB 2; Length 774;
Best Local Similarity 20.98; Pred. No. 0.00022;
Matches 121; Conservative 87; Mismatches 272; Indels 98; Gaps 25;

```
Qy 93 EAPDNVKTMLRSKGY-----FSSKVSILTEKDGAYTVHIT--PGPRTKIANVGVAILGDI 145
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

```
Db 234 EEA---LRQFYNGRYADFRVSSDAVLDESKEVYITITVDEGKKYDFGNVAV---ES 286
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 146 LSDGNLAEEYRNALENMQOPVGSDFDQDSWNSKTSVLGAVTRKGYPLAKIGNTRAANVP 205
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 287 TVPGVDGSELOGLVETRO---GASYSAKEVQOSMEAIKSRVAGEGYPPARV-TPRGDRDM 342
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 206 DTATVDLNVVDSGRPTAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDFOQALE 265
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 343 SGNTIGVTYIVDQGERAYVERIEIRGNTRDRDYVIRREFDISEGDAFNQTIITAARLRE 402
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 266 QNGHYSASVQADFDRLQG-----DRVPVKVSVTEVKRHKLETGRLDSEYGLGKIAYDY 321
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 403 ALGYFSKVNIST-----AGGSAPDRVIVVDVEQSTGSGFAGAGYSONDQVLLSEASVEE 457
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 322 YNLFNKG-YIGSVVMDMKYETTLAAGISQPRNRYGNWTSN-VSYNRSTTQNLKRAFS 379
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 458 KNFLGRGOYI-----RVAAGAGEDDARTYSLSTPEPYFLGYRLAAGDFDLFKNOSK 507
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 380 GGIWVDRDRAGIDARLGA-----EFLAEGRKIPGSDIDLGNHSHATMLTA- 423
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 508 SEDYNYDEQGFALRVTAIPITENLSTTFKYTKQINYEKGQDQWQNNANLAEPYQALIRGE 567
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 424 SWKROLLNNVLPENHGYLDGK-----IGTTLGTFLLSSTALIRTSARAGYEFT-- 471
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 568 DWTOSILSNTL---NYNLTDDRNMPREGWQAALTFNEFAGLGDSSEYKYIYAKARYYYTLS 624
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 472 PENKKLGTFIIRGOAGYTV-ARDNADVPGLMFRSGGASSVRYGELDSIGLAGPNGSVLP 530
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 625 DEYDVIGS--LTGQAGHVMPGDNLLVPDQKFE---GGRQVRGFKNDGIGPRIGSDSI-G 678
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 531 ERALLVGSLEYQLP-----FTRTLGSAVFHDMGDAAN----FKRMK-----LKHGSGL 575
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 679 GTTYFAASAETAPMPGVPEDFGLRLAG--FVDAGTMYGNKVSTSQTVKDDNSIRASAGI 736
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 576 GVRWFSPLAPSFDF-----IAYGHSDDKIRWHISLGTRF 609
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 737 GVMNASPPGPIRVDAIPIAKEDYDEEQRFRFGMSNTF 774
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

RESULT 38

```
AC1816
hypothetical protein alr0075 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1816
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriig
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1816
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-491 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB7559.1; PID:g17135053; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0075
```

Query Match 5.6%; Score 179; DB 2; Length 491;
Best Local Similarity 21.68; Pred. No. 0.00028;
Matches 106; Conservative 56; Mismatches 154; Indels 174; Gaps 20;

```
Qy 217 DSGRPIAFDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDFOQALEQNGHYSASVQ 276
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 79 DQGXPI-----IGRTQK---DFILGLLLKPKGOIFR-----EAL-----Q 111
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 277 ADFRLQ-----GDRVPKVSVEVRHKLET---GIRLDSYGLGKG 316
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 112 ADLGLQLQLESFTYVYRAYREENPFQVRI-----IYDVKERRRPPSWNYGIGNDDIGLSGR 166
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 317 IAYDYNLFNKGYIGSVVMDMKYETTLAAGISQPRNRYGNWTSNVSYNRSTT----- 370
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

Qy	415	-----NSHATMLTASAKRQL-----LNNVLHPENG-----HYLDGKIGTTLTGTLSST	457
Dd	636	SPEIKGIWDROYHTPTTSSETLDVSYONTDDYFPRNGVIFFSSYATMSGLPSSGTSLSNW	695
Qy	458	AL--IITSARAGYFFTPENKKLGCTF-----TRGOAGYTVAARDNAD--VPSCGLMFRSGG	507
Dd	696	CLGNVRNTKVYGKFAAYHHLOKYLILLDLTARKFTQCGY-IFRYNTDYLPLNSTFYMG	754
Qy	508	ASSRGVGYELDSIGUAGNGSVLP-----NGSITPKDEFGLWLGDDIFTASTELSYGVLKAAKMRLAWFPD	805
Dd	755	VTVVRGR-----RALLVGSLEYOLPSTRTLSGAVFDMDGDAANFKRMKLKGSCGLGVNFSP	805
Qy	532	-----RALLVGSLEYOLPSTRTLSGAVFDMDGDAANFKRMKLKGSCGLGVNFSP	585
Dd	806	FGFTFTPTPRGSFFYNAPTTTT---ANFKDYGVVGAGFERATWRASTGLQIEWISPMGP	861
Qy	586	--FSFDIAY-----GHSDKKIRWHISLGTRF	609
Dd	862	LVLFPFLAFENWGDKKKCKGICFNPNNDYTOHFPSMGTRF	906

omp85 analog [Imported] - Chlamydophila pneumoniae (strain J138)

A:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K. Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138
A:Reference number: AB6491; MUID:20330349; PMID:10871362
A:Accession: D86528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-790 <STO>
A:Cross-references: GB:BA000008; NID:g9878674; PIDN:BA098510.1; GSPDB:GN000
A:Experimental source: strain J138
C:Genetics:
A:Gene: *vacT*

33:

QY 22 PAADLSNKAAGFALFKNKSPDTSVKLKPFPVRIDTQDSEIKDMVEEHLPLITQQ--Q 79
- : ||| || | : : : : :
Db 87 PKVFSEGK-TNIAUHLIAKPSIRNIHISG-----NOVVPKHILKTLOIYR 132

Best Local Similarity 21.00; P-Val: 0.0010;
Matches 123; Conservative 72; Mismatches 202; Indels 188; Gaps 23;

Qy	134	I A N Y G V A L D G	: : : :	-----D I L S D G N L A	-----E Y I :
Db	189	I K Q L T F G S I R S E K S D I Q E F I O T K Q H S T T T S W F T G A G L Y H P D Y E Q D S L A I T N Y L H N N G Y	: : : :	-----D I L S D G N L A	-----E Y I :
Qy	156	R N A L E N N Q Q P V G S D F D O O D S W E N	: : : :	-----S K T S V L G A V T R K G Y P L	-----193
Db	249	A D A T	: : : :	-----V N S H Y D L D D K G N I L L Y M I D R G S R Y T L G H V H I O G F V L P K R L I E K O S Q V G	302
Qy	194	-----A K L G	: : : :	-----N P D T A T V D L N V V D S G R P I A F G D	226
Db	303	P N D L Y C P D K I W D G A H K I Q T Y A K Y G V I N T N V D V L F I P H A T R P I D Y T V Y E S G S P Y K V G L	: : : :	-----N P D T A T V D L N V V D S G R P I A F G D	362
Qy	227	F E I T G T Q R Y P E Q I V S G L A R P Q G T P Y D L D L L L D F Q Q A L E O N G H Y S G A S V	: : : :	-----Q A	277
Db	363	I K I T G N T H T K S D V L T L H E T S L F P G D T F N R L K L E D T E Q R L R N T G Y F Q S V S V Y T V R S Q L D P M G	: : : :	-----Q A	422
Qy	278	D F D R L Q C D R P V K Y S V T E	: : : :	-----V K R H K L E T G I R L D S E Y C L G C K I A Y D Y N L P N K G	328
Db	423	N A D Q Y R D I F E V K E T T T G N L G L F G F S S L N L F P G G I E L S E S N F D L F G A	: : : :	-----R N I F S K G	477
Qy	329	Y	: : : :	-----I G S V V W D M D K Y E T T L A A G I S O P N R Y N G Y T S N V	369

Db 478 FRCLRGGEHLFLKANFGDKVTDYTLKWT-----KP-HFLNTPWILGIELDKSINRAL 529
Qy 370 TONLEKRAFSGGI--WVVRD---RAGIDARLGAEEFLAEGRK-IPGSDIDILGNSHATMLTA 423
Db 530 SKDYAVQTYGGNVSTTYIILNEHLKYGLFYRGSGQTSUHERKKFLGPNIDSNKGFVSAAGV 589
Qy 424 SWKRQLLNVLHPENG-----HYLDGKIGITTLGTFLS-STALIRTSARAGYFFTPENKK 476
Db 590 NLNVDSDSPRTPTTGIRGGVTEVSGLGCTYHFTKLSLNSSIYRKLTRK----- 639
Qy 477 LGTFIIRGQAGYTVARDNA---DVPSGLMFRSGGASVRGYELDSIG----LAGPNGSVL 529
Db 640 -GILKIKGEAQFIKPYSTNTAEGVPVSERFLLGGETTVRGYKSFIIIGPKYSATEPQOGL- 697
Qy 530 PERALLVGSLEYOLPFTR--TLSGAVFHDMDAAANFKRMKLA--HGSLGVWR 579
Db 698 --SSLLI-SEEFYPLIROPNISAFVLDGFGVGLQOYKISLKLRLRSSAGFLRF 749

Search completed: November 9, 2002, 01:20:57
Job time : 28 secs

$$\frac{21^2}{3}$$

$$\frac{21^2}{43}$$